GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 03:49:53 ; Search time 5075.84 Seconds
(without alignments)
10554.310 Million cell updates/sec

Title: US-09-831-804-1_COPY_720_1955
Perfect score: 1236
Sequence: 1 atgagtgaaagtgacgaaac.....aaacatcagtgatttctcga 1236

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Search Notes

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

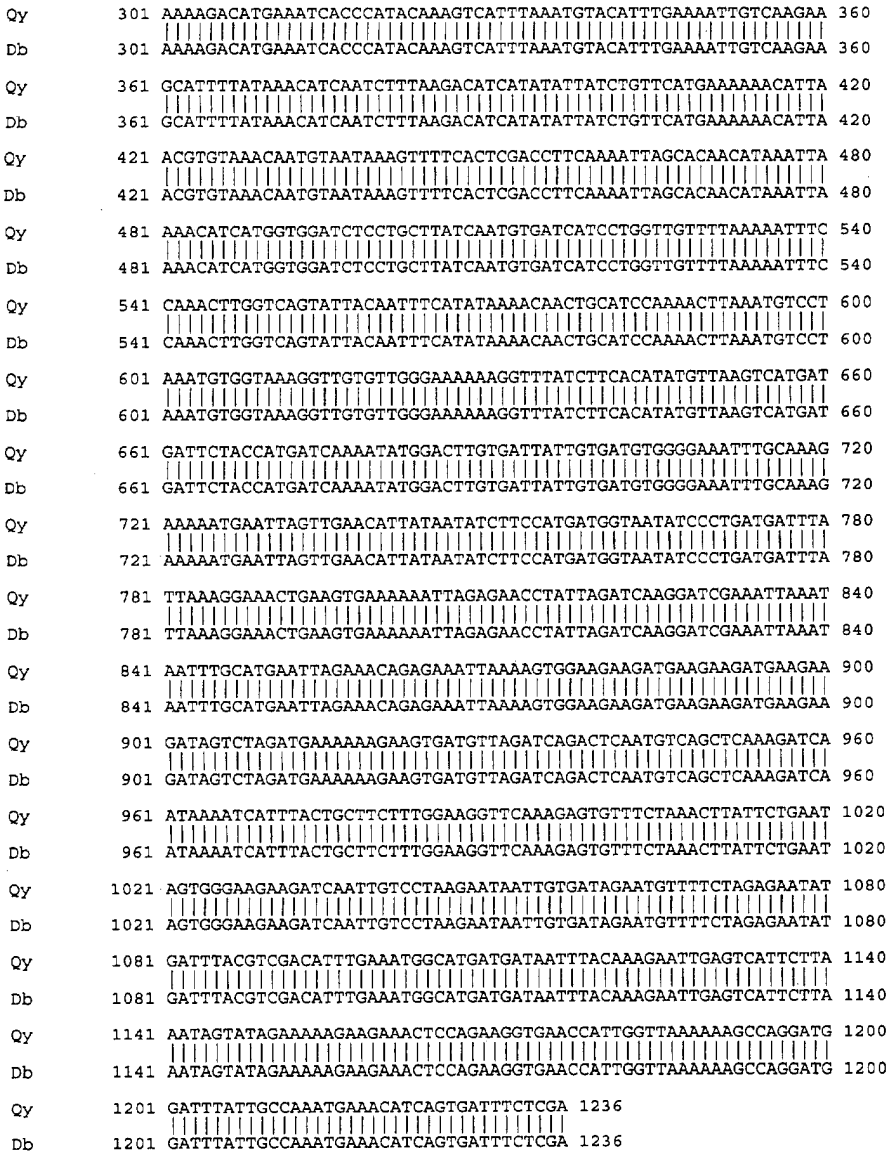
Result No.	Score	% Query Match	Length	DB	ID	Description
1	1236	100.0	1239	6	BD274350	BD274350 Candida a
2	1236	100.0	2060	6	BD274349	BD274349 Candida a
3	1232.8	99.7	1239	6	AX489193	AX489193 Sequence
4	217.4	17.6	1560	8	YSCNOFEAT	M90638 Saccharomyc
5	217.4	17.6	1739	8	YSCTFIIIA	M80611 Saccharomyc
c 6	217.4	17.6	37497	8	YSCP9677	U25841 Saccharomyc
c 7	199.2	16.1	2133	8	YSCRP026A	M33924 S.cerevisia
8	193	15.6	797	11	CNS06ING	AL400514 T7 end of
c 9	101	8.2	169063	9	AC138126	AC138126 Homo sapi
c 10	101	8.2	177299	9	AC073544	AC073544 Homo sapi
11	99.4	8.0	2132	6	AX714042	AX714042 Sequence
12	99.4	8.0	2132	9	AK056088	AK056088 Homo sapi
c 13	99.4	8.0	138627	9	AC099500	AC099500 Homo sapi
c 14	99.4	8.0	169500	2	AC138469	AC138469 Homo sapi
15	98.2	7.9	2237	9	AK122869	AK122869 Homo sapi
16	97.8	7.9	1203	9	D70831	D70831 Homo sapien
c 17	97.8	7.9	235532	9	AC008739	AC008739 Homo sapi
c 18	97.2	7.9	115995	9	AC011494	AC011494 Homo sapi
19	96.6	7.8	2320	6	AR270491	AR270491 Sequence
20	96.6	7.8	2320	9	HSU35376	U35376 Human repre
c 21	96.6	7.8	68304	10	BX000432	BX000432 Mouse DNA
22	95	7.7	981	9	HSHZF9	X78932 H.sapiens H
c 23	94.6	7.7	186233	9	AC092329	AC092329 Homo sapi
24	94.6	7.7	189317	2	AC024483	AC024483 Homo sapi
25	94.2	7.6	292390	2	AC105677	AC105677 Rattus no
26	93.6	7.6	1389	9	M27879	M27879 Homo sapien
c 27	93.6	7.6	155439	9	AC008626	AC008626 Homo sapi
28	93.4	7.6	199585	10	AC124426	AC124426 Mus muscu
c 29	92.4	7.5	41153	9	AC016628	AC016628 Homo sapi
30	92.4	7.5	158430	2	AC012431	AC012431 Homo sapi
31	92	7.4	2873	9	HUMHPLK	M55422 Human Kruep
32	92	7.4	2873	11	G28705	G28705 sWSS3976 Er
33	92	7.4	2873	11	HUMSWS1269	G18281 human chrom
34	92	7.4	156835	9	AC010620	AC010620 Homo sapi
c 35	92	7.4	191082	2	AC141066	AC141066 Homo sapi
c 36	92	7.4	203396	9	AC073210	AC073210 Homo sapi
37	91.8	7.4	1870	9	AY044432	AY044432 Homo sapi
38	91.8	7.4	2110	6	AX748253	AX748253 Sequence
39	91.8	7.4	2110	9	AK093669	AK093669 Homo sapi
40	91.8	7.4	2909	9	HSM808021	BX647875 Homo sapi
41	91.8	7.4	2976	9	BC036439	BC036439 Homo sapi
42	91.8	7.4	3078	6	AX747097	AX747097 Sequence
43	91.8	7.4	3078	9	AK091618	AK091618 Homo sapi
44	91.8	7.4	3123	9	BC040594	BC040594 Homo sapi
45	91.8	7.4	174994	9	AC022145	AC022145 Homo sapi

ALIGNMENTS

RESULT 1
BD274350
LOCUS BD274350 1239 bp DNA linear PAT 17-JUL-2003
DEFINITION Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA protein.
ACCESSION BD274350
VERSION BD274350.1 GI:33084118
KEYWORDS JP 2002531068-A/2.
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1239)
AUTHORS Pallier,F.B., Camier,S. and Sentenac,A.
TITLE Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA

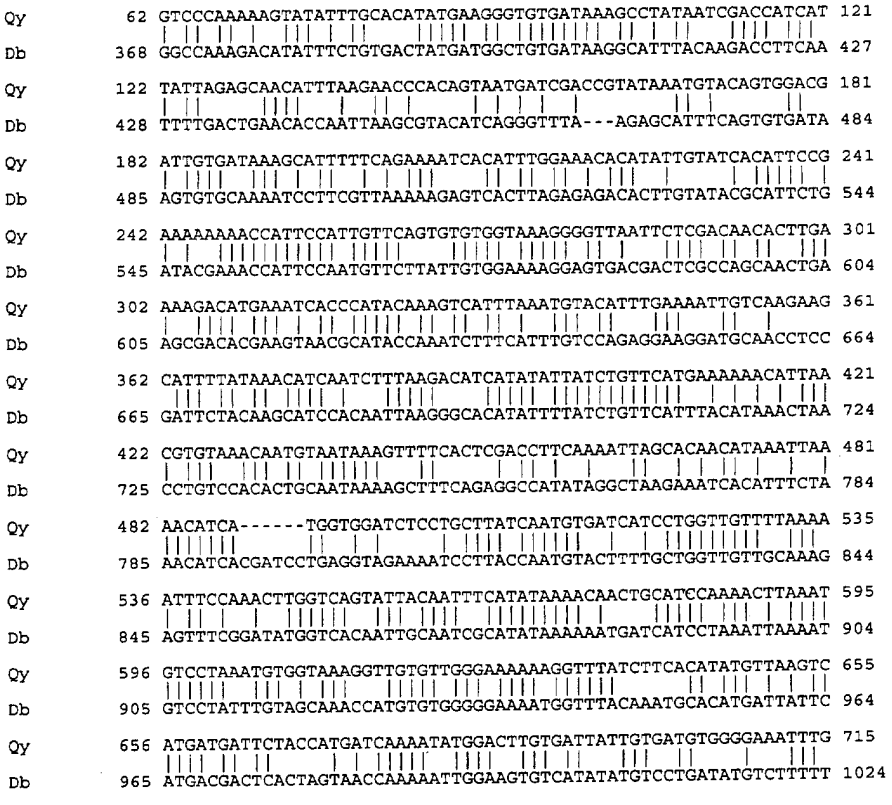
```
  JOURNAL   Patent: JP 2002531068-A 2 24-SEP-2002;
            AVENTIS PHARMA SA
COMMENT     OS   Candida albicans
            PN   JP 2002531068-A/2
            PD   24-SEP-2002
            PF   09-NOV-1999 JP 2000581204
            PR   10-NOV-1998 FR     98/14147
            PI   FLORENCE BORDON PALLIER,SYLVIE CAMIER,ANDRE SENTENAC PC
            C12N15/09,C12N15/09,A61K39/00,A61K45/00,A61P31/10,C07K14/40, PC
            C07K16/14,
            PC   C12N1/19,C12N1/21,C12P21/02,C12Q1/02,G01N33/15,G01N33/50, PC
            G01N33/53,
            PC   G01N33/566,G01N33/569//(C12N15/09,C12R1:725),C12N15/00,C12N15/
            PC   00,
            PC   (C12N15/00,C12R1:725)
            CC   Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
            CC           protein
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(1236).
FEATURES            Location/Qualifiers
     source         1. .1239
                    /organism="Candida albicans"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:5476"
ORIGIN

  Query Match           100.0%;  Score 1236;  DB 6;  Length 1239;
  Best Local Similarity  100.0%;  Pred. No. 4.3e-190;
  Matches 1236;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236
              ||||||||||||||||||||||||||||||||||||
Db       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236


RESULT 2
BD274349
LOCUS       BD274349                2060 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
            protein.
ACCESSION   BD274349
VERSION     BD274349.1  GI:33084117
KEYWORDS    JP 2002531068-A/1.
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 2060)
  AUTHORS   Pallier,F.B., Camier,S. and Sentenac,A.
  TITLE     Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
  JOURNAL   Patent: JP 2002531068-A 1 24-SEP-2002;
            AVENTIS PHARMA SA
COMMENT     OS   Candida albicans
            PN   JP 2002531068-A/1
            PD   24-SEP-2002
            PF   09-NOV-1999 JP 2000581204
            PR   10-NOV-1998 FR     98/14147
            PI   FLORENCE BORDON PALLIER,SYLVIE CAMIER,ANDRE SENTENAC PC
            C12N15/09,C12N15/09,A61K39/00,A61K45/00,A61P31/10,C07K14/40, PC
            C07K16/14,
            PC   C12N1/19,C12N1/21,C12P21/02,C12Q1/02,G01N33/15,G01N33/50, PC
            G01N33/53,
            PC   G01N33/566,G01N33/569//(C12N15/09,C12R1:725),C12N15/00,C12N15/
            PC   00,
            PC   (C12N15/00,C12R1:725)
            CC   Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
```

```
            CC              protein
            FH   Key             Location/Qualifiers
            FT   source          1. .2060
            FT                   /organism='Candida albicans'.
FEATURES             Location/Qualifiers
     source          1. .2060
                     /organism="Candida albicans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:5476"
ORIGIN

  Query Match             100.0%;  Score 1236;  DB 6;  Length 2060;
  Best Local Similarity   100.0%;  Pred. No. 3.8e-190;
  Matches 1236;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 779

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 839

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 899

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 959

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 1019

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 1079

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 1139

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 1199

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 1259

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1260 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 1319

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 1379

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1380 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 1439

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1440 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 1499

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1500 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 1559

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db       1560 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 1619

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1620 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 1679

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1739

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1799

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1859

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1919

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236
              ||||||||||||||||||||||||||||||||||||
Db       1920 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1955


RESULT 3
AX489193
LOCUS       AX489193               1239 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 6493 from Patent WO02053728.
ACCESSION   AX489193
VERSION     AX489193.1  GI:22323205
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 6493 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .1239
                     /organism="Candida albicans"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:5476"
ORIGIN

  Query Match             99.7%;  Score 1232.8;  DB 6;  Length 1239;
  Best Local Similarity   99.8%;  Pred. No. 1.4e-189;
  Matches 1234;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTATTGGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 180

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
```



RESULT 4
YSCNOFEAT
LOCUS YSCNOFEAT 1560 bp DNA linear PLN 27-APR-1993
DEFINITION Saccharomyces cerevisiae transcription factor IIIA and RNA polymerase subunit RPB6 genes.
ACCESSION M90638
VERSION M90638.1 GI:172046
KEYWORDS RNA polymerase; transcription factor IIIA.
SOURCE Saccharomyces cerevisiae (baker's yeast)
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 1560)
AUTHORS Woychik,N.A. and Young,R.A.
TITLE Genes encoding transcription factor IIIA and the RNA polymerase common subunit RPB6 are divergently transcribed in Saccharomyces cerevisiae
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (9), 3999-4003 (1992)
MEDLINE 92237295
PUBMED 1570325
COMMENT Original source text: Saccharomyces cerevisiae (strain S288C) DNA.
FEATURES Location/Qualifiers
source 1. .1560
/organism="Saccharomyces cerevisiae"
/mol_type="genomic DNA"
/strain="S288C"
/db_xref="taxon:4932"
ORIGIN chromosome 14.

Query Match 17.6%; Score 217.4; DB 8; Length 1560;
Best Local Similarity 59.4%; Pred. No. 9.9e-26;
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;



```
RESULT 5
YSCTFIIIA
LOCUS       YSCTFIIIA               1739 bp    DNA     linear   PLN 13-SEP-1996
DEFINITION  Saccharomyces cerevisiae transcription factor IIIA (TFIIIA) gene,
            complete cds.
ACCESSION   M80611
VERSION     M80611.1  GI:172902
KEYWORDS    transcription factor IIIA; zinc-finger protein; zinc-finger
            transcription factor.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1739)
  AUTHORS   Archambault,J., Milne,C.A., Schappert,K.T., Baum,B., Friesen,J.D.
            and Segall,J.
  TITLE     The deduced sequence of the transcription factor TFIIIA from
            Saccharomyces cerevisiae reveals extensive divergence from Xenopus
            TFIIIA
  JOURNAL   J. Biol. Chem. 267 (5), 3282-3288 (1992)
  MEDLINE   92147684
   PUBMED   1737784
FEATURES             Location/Qualifiers
     source          1. .1739
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="W303-1A"
                     /db_xref="taxon:4932"
                     /chromosome="XVI"
     gene            234. .1523
                     /gene="TFIIIA"
     CDS             234. .1523
                     /gene="TFIIIA"
                     /note="zinc finger containing protein"
                     /codon_start=1
                     /product="transcription factor IIIA"
                     /protein_id="AAB08014.1"
                     /db_xref="GI:172903"
                     /translation="MGGEVLNNEGMPLAELKQETIPISRSESSESLNSLTSTRSSSSN
                     RPKTYFCDYDGCDKAFTRPSILTEHQLSVHQGLRAFQCDKCAKSFVKKSHLERHLYTH
                     SDTKPFQCSYCGKGVTTRQQLKRHEVTHTKSFICPEEGCNLRFYKHPQLRAHILSVHL
                     HKLTCPHCNKSFQRPYRLRNHISKHHDPEVENPYQCTFAGCCKEFRIWSQLQSHIKND
                     HPKLKCPICSKPCVGENGLQMHMIIHDDSLVTKNWKCHICPDMSFSRKHDLLTHYGSI
                     HTEEDIPLELKYKISDIQQLVQDHGVQLGNSKHSNEQDEEKISNRLRKRRKLTENNNV
                     EFLQNEVDLEKRLESGENGLNLLLNTVGRKYRCFYNNCSRTFKTKEKYEKHIDKHKVH
                     ELKLKILQEKEENKTLVDQNHKEPFIIQKETQSAGDK"
ORIGIN

  Query Match           17.6%;  Score 217.4;  DB 8;  Length 1739;
  Best Local Similarity   59.4%;  Pred. No. 9.6e-26;
  Matches  409;  Conservative    0;  Mismatches  271;  Indels    9;  Gaps    2;

Qy         62 GTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCAT 121
              | || || |  ||| | ||    ||||| || |||||||| || | ||   |||| |||
Db        367 GGCCAAAGACATATTTCTGTGACTATGATGGCTGTGATAAGGCATTTACAAGACCTTCAA 426

Qy        122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACG 181
              | ||     ||||   | || |     | | | | |         || |       ||
Db        427 TTTTGACTGAACACCAATTAAGCGTACATCAGGGTTTA---AGAGCATTTCAGTGTGATA 483

Qy        182 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCG 241
              | |||| ||| | || | | |||    || || || | |||  |     | |||||| |
Db        484 AGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTATACGCATTCTG 543

Qy        242 AAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGA 301
              | |  |||||||||||| |||||   ||||| ||||| || |   |||| || ||  |||
Db        544 ATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCGCCAGCAACTGA 603

Qy        302 AAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAG 361
              |  |||| ||| | || ||||| || || || |  ||| ||   |||   ||  |
Db        604 AGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGGATGCAACCTCC 663

Qy        362 CATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAA 421
              ||| || || |||| |   |||||   ||||| ||||||||||||  | | | | |||
Db        664 GATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTTACATAAACTAA 723

Qy        422 CGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAA 481
              | |||  ||| || ||||||    ||      | || |  |   ||  | | ||  | |
Db        724 CCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAATCACATTTCTA 783

Qy        482 AACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAA 535
              |||||||      ||  | |     | |||| ||||||   |   | |||||||||  |||
Db        784 AACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGGTTGTTGCAAAG 843

Qy        536 ATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAAT 595
              | ||  |  ||||||  ||| ||||  |||||||| |     |||| ||| | ||||
Db        844 AGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCCTAAATTAAAAT 903

Qy        596 GTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTC 655
              ||||||  ||| | |||  ||||| ||| |||| |||||||      || ||| | | ||
Db        904 GTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCACATGATTATTC 963

Qy        656 ATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTG 715
              |||| || ||     | | |||||  |||| ||| ||   |||  ||    |  |||
Db        964 ATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGATATGTCTTTTT 1023

Qy        716 CAAAGAAAAATGAATTAGTTGAACATTAT 744
              | |  ||| ||||  |  |    ||||||
Db       1024 CTAGAAAACATGATCTTCTCACGCATTAT 1052


RESULT 6
YSCP9677/c
LOCUS       YSCP9677                37497 bp    DNA     linear   PLN 01-AUG-1997
DEFINITION  Saccharomyces cerevisiae chromosome XVI cosmid 9677.
ACCESSION   U25841 U00094
VERSION     U25841.1  GI:786295
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 37497)
  AUTHORS   Johnston,M., Andrews,S., Brinkman,R., Cooper,J., Ding,H., Du,Z.,
            Favello,A., Fulton,L., Gattung,S., Greco,T., Kirsten,J., Kucaba,T.,
            Hallsworth,K., Hawkins,J., Hillier,L., Jier,M., Johnson,D.,
            Johnston,L., Langston,Y., Latreille,P., Le,T., Mardis,E.,
            Menezes,S., Miller,N., Nhan,M., Pauley,A., Peluso,D., Rifken,L.,
            Riles,L., Taich,A., Trevaskis,E., Vignati,D., Wilcox,L.,
            Wohldman,P., Vaudin,M., Wilson,R. and Waterston,R.
  TITLE     The sequence of a portion of the right arm of Saccharomyces
            cerevisiae chromosome XVI
  JOURNAL   Unpublished (1995)
REFERENCE   2  (bases 1 to 37497)
  AUTHORS   Miller,N.
  TITLE     The sequence of S. cerevisiae cosmid 9677
  JOURNAL   Unpublished (1995)
REFERENCE   3  (bases 1 to 37497)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-1995) Robert Waterston
COMMENT     Submitted by:
                     Genome Sequencing Center
                     Department of Genetics, Washington University,
                     St. Louis, MO  63110, USA
                     e-mail: mj@sequencer.wustl.edu

            NEIGHBORING COSMID INFORMATION:
            This sequence includes nucleotides 1-37497 of cosmid 9677.  The
            cosmid on the left is 9705.
FEATURES             Location/Qualifiers
     source          1. .37497
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
```

/strain="S288C (AB972)"
/db_xref="taxon:4932"
/map="XVI"
/note="baker's yeast"
gene complement(2366. .3655)
/gene="TFC2"
CDS complement(2366. .3655)
/gene="TFC2"
/note="P9677.9"
/codon_start=1
/product="Transcription factor TFIIIA (PIR accession number S20050)"
/protein_id="AAB64615.1"
/db_xref="GI:786305"
/translation="MGGEVLNNEGMPLAELKQETIPISRSESSESLNSLTSTRSSSSN
RPKTYFCDYDGCDKAFTRPSILTEHQLSVHQGLRAFQCDKCAKSFVKKSHLERHLYTH
SDTKPFQCSYCGKGVTTRQQLKRHEVTHTKSFICPEEGCNLRFYKHPQLRAHILSVHL
HKLTCPHCNKSFQRPYRLRNHISKHHDPEVENPYQCTFAGCCKEFRIWSQLQSHIKND
HPKLKCPICSKPCVGENGLQMHMIIHDDSLVTKNWKCHICPDMSFSRKHDLLTHYGSI
HTEEDIPLELKYKISDIQQLVQDHGVQLGNSKHSNEQDEEKISNRLRKRRKLTENNNV
EFLQNEVDLEKRLESGENGLNLLLNTVGRKYRCFYNNCSRTFKTKEKYEKHIDKHKVH
ELKLKILQEKEENKTLVDQNHKEPFIIQKETQSAGDK"
gene 3890. .4433
/gene="RPB6"
CDS join(3890. .3909,3986. .4433)
/gene="RPB6"
/note="P9677.8"
/codon_start=1
/product="23 kD polypeptide of RNA Polymerases I, II, and III (Swiss Prot. accession number P20435)"
/protein_id="AAB64616.1"
/db_xref="GI:786304"
/translation="MSDYEEAFNDGNENFEDFDVEHFSDEETYEEKPQFKDGETTDAN
GKTIVTGGNGPEDFQQHEQIRRKTLKEKAIPKDQRATTPYMTKYERARILGTRALQIS
MNAPVFVDLEGETDPLRIAMKELAEKKIPLVIRRYLPDGSFEDWSVEELIVDL"
gene complement(4621. .5112)
/gene="P9677.10"
CDS complement(4621. .5112)
/gene="P9677.10"
/note="Similar to calmodulins (e.g., Swiss Prot. accession number P13868) and calmodulin-related proteins (e.g., Swiss Prot. accession number P27482) from several organism."
/codon_start=1
/evidence=not_experimental
/protein_id="AAB64617.1"
/db_xref="GI:786306"
/translation="MDHSESLTFNQLTQDYINKLKDAFQMLDEDEDGLISRGDLTKIY
ATLGKTLTDEEWSKMVPDNDTSTAEVGEEGVSFPIFLSIMGKNLSQFPEREELEESLK
AIGRGHDLNVPLNEVIDSLKEAGFENPEEEFAKLFKLFTTNQQATEERTFRGKLFLDS
ITD"
gene 5297. .9595
/gene="SKI3"
CDS 5297. .9595
/gene="SKI3"
/note="P9677.7"
/codon_start=1
/product="Contains 8 copies of the TPR domain"
/protein_id="AAB64618.1"
/db_xref="GI:786303"
/translation="MSDIKQLLKEAKQELTNRDYEETIEISEKVLKLDPDNYFAHIFL
GKALSSLPASNNVSSNRNLERATNHYVSAAKLVPDNLLAWKGLFLLFRTTEVVPDILS
YDEYFDLCGQYADALLKQEQSQVELINDIKLLKKTHPDCQKAFYQHLKPGSLMAETIG
RHLSTPQDALLNLIKILSNIETTEIGKTLSQNRLKLKASDPDYQIKLNSFSWEIIKNS
EIDQLYNQLVNILADDQKRSEIENQWLEYRIKVLKSMPLDVKKDFFTKVKEMVEDMVL
VNHQSLLAWQKYFEWTDYEDLDNMDAPLIIKYFKKFPKDPLAMILYSWLSSKLSKYDI
KSLESANKPPEGHKKTEKETDIKDVDETNEDEVKDRVEDEVKDRVEDEVKDQDEEAKE
DEEEDLDDIEIGLLEEEVVTVLTENIVKCKNNILAHRILCQYYLLTKEYEAALPYIKN
GISLIAYNIKDLGVHLPLTKREFSLDLATVYTYVDAPKDHNAALKLYDNILSGDFSNI
QAKMGKGIIFIERKNWKDAMTLLTQVHEQSPNNLEVLSELSWSKAHMGYMDEALAGLD
TVIKGIKGMDLRSIDFRALNLWRQAKVYIMKHASINDAKQENVKCAFKLLIQSIKILD
TFAPGFSTLGDIYCHYYKDHLRAFKCYFKAFDLDAGDYTAAKYITETYASKPNWQAAS
SIASRLIKGEKAKAELRSNNWPFRVVGIAHLEKQEESDSIEWFQSALRVDPNDVESWV
GLGQAYHACGRIEASIKVFDKAIQLRPSHTFAQYFKAISLCDVGEYLESLDILEKVCQ
EAATEESFQIGLVEVLMRCSLDLYSQGFLLKSVSIAKDTIERIKIIISELKCENQQVW
IYLSQVLRLFIWIESKVDTLPVESLVSIFENSQFSGSEEIDSVDNIKIDTLLDSTTDD
NVSIACKFLILASKYSVSDQKFTDIAGTVRASYWYNIGISELTAFITLKEPQYRDAAI
FAFKKSIQLQSNTSETWIGLGIATMDINFRVSQHCFIKATALEPKATNTWFNLAMLGL
KKKDTEFAQQVLNKLQSLAPQDSSPWLGMALILEEQGDIIGSSKLFAHSFILSNGRSK
AAQFMYAKNVLENHINNGDDERDIETVEKLTTASIALEQFFKKSPDSQFALQCALLTL
ERLHHYENANELANRLIGILEKKFEKTQDERELFNFAIIKGQFARIHLGLGNFELSIE
NADLSQGIISESSDEKSMKTKISNHICLGLSYFFLNDFDQTLNQFQELLSISKDSKHL
VVLIAKVLYDVGESDTKEIALQELTEYIATSGADLLVTLTIAAMSILDDKREDLSIIL
EELKALPLSKQIIDKHKDAPYLIEEITKRLYRNDTGKQVWQRSAYFFPNNLKVWERLD
KNIQRRIASNGQNKVTAEEMSKLYCESKNLRSIQRGMFLCPWNVTAVKALNECF"
gene complement(9710. .11674)
/gene="RPC3"
CDS complement(9710. .11674)
/gene="RPC3"
/note="P9677.11"
/codon_start=1
/product="74 kD subunit of RNA Polymerase III (Swiss Prot. accession number P32349)"
/protein_id="AAB64619.1"
/db_xref="GI:786307"
/translation="MDELLGEALSAENQTGESTVESEKLVTPEDVMTISSLEQRTLNP
DLFLYKELVKAHLGERAASVIGMLVALGRLSVRELVEKIDGMDVDSVKTTLVSLTQLR
CVKYLQETAISGKKTTYYYYNEEGIHILLYSGLIIDEIITQMRVNDEEEHKQLVAEIV
QNVISLGSLTVEDYLSSVTSDSMKYTISSLFVQLCEMGYLIQISKLHYTPIEDLWQFL
YEKHYKNIPRNSPLSDLKKRSQAKMNAKTDFAKIINKPNELSQILTVDPKTSLRIVKP
TVSLTINLDRFMKGRRSKQLINLAKTRVGSVTAQVYKIALRLTEQKSPKIRDPLTQTG
LLQDLEEAKSFQDEAELVEEKTPGLTFNAIDLARHLPAELDLRGSLLSRKPSDNKKRS
GSNAAASLPSKKLKTEDGFVIPALPAAVSKSLQESGDTQEEDEEEEDLDADTEDPHSA
SLINSHLKILASSNFPFLNETKPGVYYVPYSKLMPVLKSSVYEYVIASTLGPSAMRLS
RCIRDNKLVSEKIINSTALMKEKDIRSTLASLIRYNSVEIQEVPRTADRSASRAVFLF
RCKETHSYNFMRQNLEWNMANLLFKKEKLKQENSTLLKKANRDDVKGRENELLLPSEL
NQLKMVNERELNVFARLSRLLSLWEVFQMA"
gene 12014. .13120
/gene="QCR2"
CDS 12014. .13120
/gene="QCR2"
/note="P9677.6"
/codon_start=1
/product="Ubiquinol-cytochrome C reductase core protein 2"
/protein_id="AAB64620.1"
/db_xref="GI:786302"
/translation="MLSAARLQFAQGSVRRLTVSARDAPTKISTLAVKVHGGSRYATK
DGVAHLLNRFNFQNTNTRSALKLVRESELLGGTFKSTLDREYITLKATFLKDDLPYYV
NALADVLYKTAFKPHELTESVLPAARYDYAVAEQCPVKSAEDQLYAITFRKGLGNPLL
YDGVERVSLQDIKDFADKVYTKENLEVSGENVVEADLKRFVDESLLSTLPAGKSLVSK
SEPKFFLGEENRVRFIGDSVAAIGIPVNKASLAQYEVLANYLTSALSELSGLISSAKL
DKFTDGGLFTLFVRDQDSAVVSSNIKKIVADLKKGKDLSPAINYTKLKNAVQNESVSS
PIELNFDAVKDFKLGKFNYVAVGDVSNLPYLDEL"
gene 14493. .15410
/gene="P9677.5"
CDS 14493. .15410
/gene="P9677.5"
/note="Belongs to the MIP family of transmembrane channel proteins."
/codon_start=1
/evidence=not_experimental
/protein_id="AAB64621.1"
/db_xref="GI:786301"
/translation="MSSNDSNDTDKQHTRLDPTGVDDAYIPPEQPETKHHRFKISRDT
LRDHFIAAVGEFCGTFMFLWCAYVICNVANHDVALVAAPDGSHPGQLIMIAIGFGFSV
MFSIWCFAGVSGGALNPAMSLSLCLARAVSPTRCVVMWVSQIVAGMAAGGAASAMTPG
EVLFANSLGLGCSRTRGLFLEMFGTAILCLTVLMTAVEKRETNFMAALPIGISLFIAH
VALTAYTGTGVNPARSLGAAVAARYFPHYHWIYWIGTLLGSILAWSVWQLLQILDYTT
YVTAEKAASTKEKAQKKGETSSSSAVAEV"
gene complement(15542. .16012)
/gene="P9677.12"
CDS complement(15542. .16012)
/gene="P9677.12"
/note="Similar in C-terminus to Gentamicin 3'-acetyltransferase of P. aeruginosa (Swiss Prot. accession number P23181)"

```
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB64622.1"
                     /db_xref="GI:786308"
                     /translation="MSNTSEDNITVRFVTENDKEGWQRLWKSYQDFYEVSFPDDLDDF
                     NFGRFLDPNIKMWAAVAVESSSEKIIGMINFFNHMTTWDFKDKIYINDLYVDENSRVK
                     GAGGKLIQFVYDEADKLGTPSVYWCTDESNHRAQLLYVKVGYKAPKILYKRKGY"
     gene            complement(16937. .19570)
                     /gene="P9677.13"
     CDS             complement(16937. .19570)
                     /gene="P9677.13"
                     /note="Similar to S. cerevisiae hypothetical protein
                     HRD799 (PIR accession number S45161) and S. pombe
```

Query Match 17.6%; Score 217.4; DB 8; Length 37497;
Best Local Similarity 59.4%; Pred. No. 4.7e-26;
Matches 409; Conservative 0; Mismatches 271; Indels 9; Gaps 2;

```
Qy         62 GTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCAT 121
Db       3522 GGCCAAAGACATATTTCTGTGACTATGATGGCTGTGATAAGGCATTTACAAGACCTTCAA 3463

Qy        122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACG 181
Db       3462 TTTTGACTGAACACCAATTAAGCGTACATCAGGGTTTA---AGAGCATTTCAGTGTGATA 3406

Qy        182 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCG 241
Db       3405 AGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTATACGCATTCTG 3346

Qy        242 AAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGA 301
Db       3345 ATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCGCCAGCAACTGA 3286

Qy        302 AAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAG 361
Db       3285 AGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGGATGCAACCTCC 3226

Qy        362 CATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAA 421
Db       3225 GATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTTACATAAACTAA 3166

Qy        422 CGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAA 481
Db       3165 CCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAATCACATTTCTA 3106

Qy        482 AACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAA 535
Db       3105 AACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGGTTGTTGCAAAG 3046

Qy        536 ATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAAT 595
Db       3045 AGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCCTAAATTAAAAT 2986

Qy        596 GTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTC 655
Db       2985 GTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCACATGATTATTC 2926

Qy        656 ATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTG 715
Db       2925 ATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGATATGTCTTTTT 2866

Qy        716 CAAAGAAAAATGAATTAGTTGAACATTAT 744
Db       2865 CTAGAAAACATGATCTTCTCACGCATTAT 2837
```

```
RESULT 7
YSCRP026A/c
LOCUS       YSCRP026A               2133 bp    DNA     linear   PLN 27-APR-1993
DEFINITION  S.cerevisiae RNA polymerase II sixth subunit (RP026) gene, complete
            cds.
ACCESSION   M33924
VERSION     M33924.1  GI:172452
KEYWORDS    RNA polymerase II; c-myc proto-oncogene; transmembrane protein;
            tyrosine kinase.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2133)
  AUTHORS   Archambault,J., Schappert,K.T. and Friesen,J.D.
  TITLE     A suppressor of an RNA polymerase II mutation of Saccharomyces
            cerevisiae encodes a subunit common to RNA polymerases I, II, and
            III
  JOURNAL   Mol. Cell. Biol. 10 (12), 6123-6131 (1990)
  MEDLINE   91061718
   PUBMED   2247052
COMMENT     Original source text: S.cerevisiae single-copy plasmid DNA, clone
            2.
            Draft entry and computer-readable sequence for [Unpublished (1990)]
            kindly submitted
            by J.D.Friesen, 01-MAY-1990.
               The Hospital for Sick Children
               555 University Avenue
               Toronto, Ontario, CANADA M5G 1X8
               e-mail: jim@sickkids.toronto.edu.
FEATURES             Location/Qualifiers
     source          1. .2133
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
     CDS             join(1026. .1045,1122. .1569)
                     /note="RNA polymerase II sixth subunit (RP026)"
                     /codon_start=1
                     /protein_id="AAA34989.1"
                     /db_xref="GI:172453"
                     /translation="MSDYEEAFNDGNENFEDFDVEHFSDEETYEEKPQFKDGETTDAN
                     GKTIVTGGNGPEDFQQHEQIRRKTLKEKAIPKDQRATTPYMTKYERARILGTRALQIS
                     MNAPVFVDLEGETDPLRIAMKELAEKKIPLVIRRYLPDGSFEDWSVEELIVDL"
     exon            <1026. .1045
                     /note="RNA polymerase II sixth subunit (RP026)"
                     /number=1
     intron          1046. .1121
                     /note="RP026 intron A"
     exon            1122. .>1569
                     /note="RNA polymerase II sixth subunit (RP026)"
                     /number=2
ORIGIN      Chromosome XVI.
```

Query Match 16.1%; Score 199.2; DB 8; Length 2133;
Best Local Similarity 61.4%; Pred. No. 8e-23;
Matches 340; Conservative 0; Mismatches 208; Indels 6; Gaps 1;

```
Qy        171 TACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGT 230
Db        556 TTCAGTGTGATAGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTA 497

Qy        231 ATCACATTCCGAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCG 290
Db        496 TACGCATTCTGATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCG 437

Qy        291 ACAACACTTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAA 350
Db        436 CCAGCAACTGAAGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGG 377

Qy        351 TTGTCAAGAAGCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGA 410
Db        376 ATGCAACCTCCGATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTT 317

Qy        411 AAAAACATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACA 470
Db        316 ACATAAACTAACCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAA 257

Qy        471 ACATAAATTAAAACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGG 524
```

```
Db        256 TCACATTTCTAAACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGG 197

Qy        525 TTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCC 584

Db        196 TTGTTGCAAAGAGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCC 137

Qy        585 AAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACA 644

Db        136 TAAATTAAAATGTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCA 77

Qy        645 TATGTTAAGTCATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGT 704

Db         76 CATGATTATTCATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGA 17

Qy        705 GGGGAAATTTGCAA 718

Db         16 TATGTCTTTTTCTA 3
```

RESULT 8
CNS06ING
LOCUS CNS06ING 797 bp DNA linear STS 30-NOV-2001
DEFINITION T7 end of clone AS0AA022B07 of library AS0AA from strain CLIB 533 of Saccharomyces bayanus, sequence tagged site.
ACCESSION AL400514
VERSION AL400514.1 GI:12156635
KEYWORDS STS.
SOURCE Saccharomyces bayanus
ORGANISM Saccharomyces bayanus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 797)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 797)
AUTHORS Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P., Aigle,M. and Durrens,P.
TITLE Genomic exploration of the hemiascomycetous yeasts: 5. Saccharomyces bayanus var. uvarum
JOURNAL FEBS Lett. 487 (1), 37-41 (2000)
MEDLINE 20584715
PUBMED 11152880
REFERENCE 3 (bases 1 to 797)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This STS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES Location/Qualifiers
source 1. .797
/organism="Saccharomyces bayanus"
/mol_type="genomic DNA"
/strain="CLIB 533"
/variety="uvarum"
/db_xref="taxon:4931"
/clone="AS0AA022B07"
/clone_lib="AS0AA"
/note="end : T7"
misc_feature <2. .>796
/note="similar to Saccharomyces cerevisiae ORF YPR186c [PZF1 ; TFIIIA (transcription initiation factor)]"
/evidence=not_experimental
ORIGIN

Query Match 15.6%; Score 193; DB 11; Length 797;
Best Local Similarity 59.3%; Pred. No. 1e-21;
Matches 347; Conservative 1; Mismatches 231; Indels 6; Gaps 1;

```
Qy        178 GACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACAT 237

Db         83 GACAATTGTGCAAAATCCTTTGTTAARAAGAGTCATCTAGAGAGGCACTTGTTTTCGCAT 142

Qy        238 TCCGAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACAC 297

Db        143 TCTGATACGAAGCCATTCCAATGTTCTTATTGTGGCAAGGGGGTGACAACTCGACAGCAA 202

Qy        298 TTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAA 357

Db        203 TTGAAGCGCCACGAAGTTACACATACCAAATCTTTTGTTTGTCCGGAAGATGGATGTGAT 262

Qy        358 GAAGCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACA 417

Db        263 CTCCGATTTTATAAGCACCCTCAATTAAGGGCTCATATTCTATCTGTCCACCTAGACAAA 322

Qy        418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477

Db        323 TTAACCTGTCCACAATGTAGTAAAARTTTTCAGAGACCCTACAGGCTAAGGAATCATATC 382

Qy        478 TTAAAACATCATGGT------GGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTT 531

Db        383 TCTAAACATCATGATCCTGAAGTGGTAAATCCGTATCAATGTACTTTCGCTGGCTGTTGT 442

Qy        532 AAAAATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTT 591

Db        443 ACAGAGTACCGTATATGGTCACAGTTGCAATCACATATCAAAAATGACCATCCTAAATTG 502

Qy        592 AAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTA 651

Db        503 AAATGCCCCATTTGCAACAAGCCTTGTGTAGGGGAAAATGGTCTACAAATGCATATGATC 562

Qy        652 AGTCATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAA 711

Db        563 ATTCATGATGAATCTTTGGTAACTAAGAACTGGAAATGTCATATTTGTTCTAATACATCT 622

Qy        712 TTTGCAAAGAAAAATGAATTAGTTGAACATTATAATATCTTCCAT 756

Db        623 TTCTTGAGAAAGCACGACCTCCTTACTCACTACCAGATGACGCAT 667
```

RESULT 9
AC138126/c
LOCUS AC138126 169063 bp DNA linear PRI 17-DEC-2002
DEFINITION Homo sapiens chromosome 19 clone RP11-274A19, complete sequence.
ACCESSION AC138126
VERSION AC138126.1 GI:27151357
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169063)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 169063)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission

```
  JOURNAL   Submitted (17-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1..169063
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="RP11-274A19"
ORIGIN

  Query Match             8.2%;  Score 101;  DB 9;  Length 169063;
  Best Local Similarity   50.8%;  Pred. No. 2e-07;
  Matches  298;  Conservative    0;  Mismatches  280;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
Db      80134 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 80075

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
Db      80074 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 80015

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
Db      80014 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 79961

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
Db      79960 GGCAAAGCCTTTAACCATTCCTCACATCTTGCTACACATAAGGTAATTCATACTGGAGAG 79901

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
Db      79900 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 79841

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
Db      79840 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 79781

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
Db      79780 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 79721

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
Db      79720 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 79661

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
Db      79660 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 79601

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
Db      79600 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTACAAGT 79554


RESULT 10
AC073544/c
LOCUS       AC073544               177299 bp    DNA     linear   PRI 19-JUN-2002
DEFINITION  Homo sapiens chromosome 19 clone RP11-359H18, complete sequence.
ACCESSION   AC073544
VERSION     AC073544.4  GI:21465367
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177299)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177299)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 177299)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUN-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jun 19, 2002 this sequence version replaced gi:13699752.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.2.
            NOTE:  BACTERIAL TRANSPOSONS excised at 14620 and 98099.
FEATURES             Location/Qualifiers
     source          1..177299
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="RP11-359H18"
ORIGIN

  Query Match             8.2%;  Score 101;  DB 9;  Length 177299;
  Best Local Similarity   50.8%;  Pred. No. 2e-07;
  Matches  298;  Conservative    0;  Mismatches  280;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
Db     163740 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 163681

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
Db     163680 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 163621

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
Db     163620 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 163567

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
Db     163566 GGCAAAGCCTTTAACCATTCCTCACATCTTGCTACACATAAGGTAATTCATACTGGAGAG 163507

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
Db     163506 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 163447

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
Db     163446 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 163387

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
Db     163386 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 163327

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
Db     163326 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 163267

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
Db     163266 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 163207

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
```

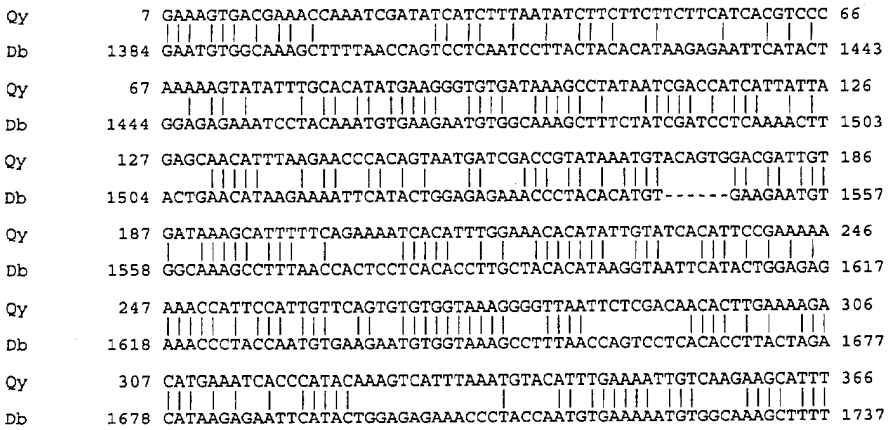
```
Db     163206 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTACAAGT 163160
```

```
RESULT 11
AX714042
LOCUS       AX714042               2132 bp    DNA     linear   PAT 15-APR-2003
DEFINITION  Sequence 726 from Patent EP1293569.
ACCESSION   AX714042
VERSION     AX714042.1  GI:29888970
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
            Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
            Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
            Masuho,Y.
  TITLE     Full-length cDNAs
  JOURNAL   Patent: EP 1293569-A 726 19-MAR-2003;
            Helix Research Institute (JP) ; Research Association for
            Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1. .2132
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

```
  Query Match             8.0%;  Score 99.4;  DB 6;  Length 2132;
  Best Local Similarity   50.6%;  Pred. No. 1e-06;
  Matches  297;  Conservative    0;  Mismatches  281;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |          || || | | || | |   |   |   | | | |
Db       1384 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 1443

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |   | || || |||||  ||||  ||||| |   |||||  | |||  | |
Db       1444 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 1503

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
               |||||   | ||  || | |   ||   ||| || | ||||       || || |||
Db       1504 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 1557

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| ||| |       |||||  | |  ||||||  |||   ||| | | | |
Db       1558 GGCAAAGCCTTTAACCACTCCTCACACCTTGCTACACATAAGGTAATTCATACTGGAGAG 1617

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | ||| |||   ||  ||||||||||   ||||         |||| | |  |||
Db       1618 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 1677

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |  | |||||                |   || |||||| |||    |||| |||
Db       1678 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 1737

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
               |  |       |  | |  |||| | | | |   |  | ||| |||   | | |
Db       1738 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 1797

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||||  |||||| |  | |||          ||||||| || |  ||||||||  |||
Db       1798 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 1857

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
               ||  |||||      ||  ||     |||  | |||     | | | || |
Db       1858 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 1917

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
              ||||| |       ||  ||     |||  | |||     | | | || |
Db       1918 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTATAAGT 1964
```

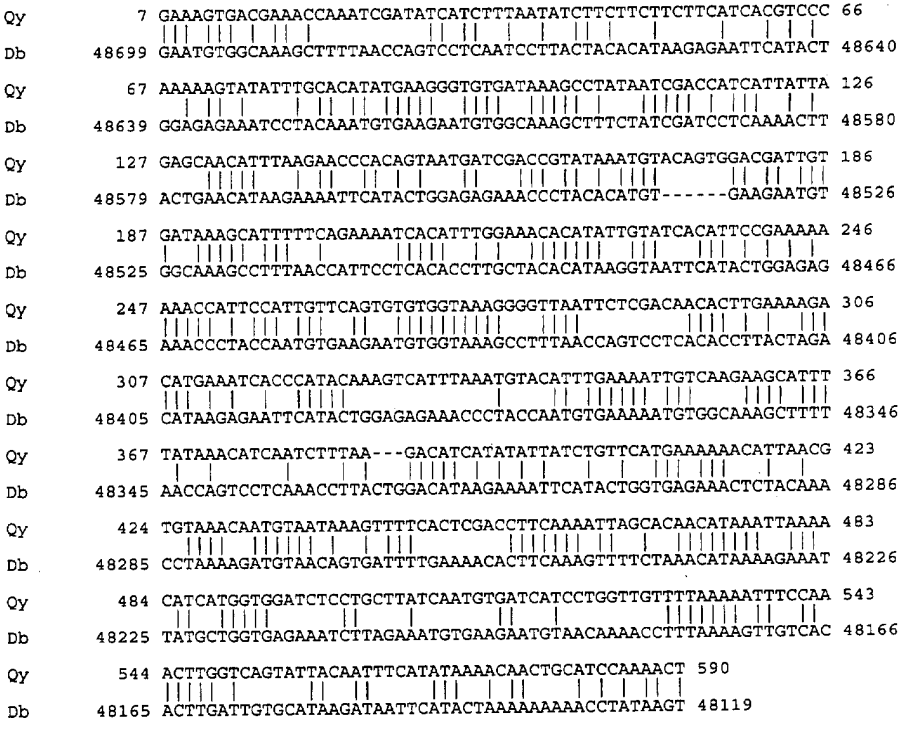
```
RESULT 12
AK056088
LOCUS       AK056088               2132 bp    mRNA    linear   PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ31526 fis, clone NT2RI2000341, moderately
            similar to ZINC FINGER PROTEIN 43.
ACCESSION   AK056088
VERSION     AK056088.1  GI:16551397
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
            Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,
            Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
            Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
            Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2132)
  AUTHORS   Isogai,T., Otsuki,T. and Sugiyama,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: RAB and
            HRI.
FEATURES             Location/Qualifiers
     source          1. .2132
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RI2000341"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone_lib="NT2RI2"
                     /note="cloning vector: pME18SFL3~mRNA from NT2 neuronal
                     precursor cells treated 2-weeks mitotic inhibitor after
                     5-weeks retinoic acid (RA) induction.~majorly NT2 neuron"
     CDS             148. .1878
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB71090.1"
                     /db_xref="GI:16551398"
                     /translation="MVAEPPVICSHFAQHFSPEQNIKDSFQKVTPRRYGKCEHENLQL
                     SKSVDECKVQKGGYNGLNQCLPTTQSKIFQCDKYMKIFHKFSNLNGHKVRHTRKKPFK
                     YKEFGKSFCIFSNLTQHKIICTRVNFYKCEDCGKAFNGSSIFTKHKRIHIGEKSYICE
                     ECGKACNQFTNLTTHKIIYTRDKLYKREECSKAFNLSSHITTHTIIHTGENPYKREEC
                     DKAFNQSSTLTTHKIIHTREKLNEYKECGKAFNQSSHLTRHKIIHTGEKPYKCEECGK
                     AFNQSSHLTRHKIIHTGEKPYKCEECGKAFRQSSHLTTHKIIHTGEKPYKCEECGKAF
                     NKSSHLTRHKSIHTGEKPYQCEKCGKASNQSSNLTEHKNIHTEEKPYKCEECGKAFNQ
                     FSNLTTHKRIHTGEKPYKCEECGKAFNQSSILTTHKRIHTGEKSYKCEECGKAFYRSS
                     KLTEHKKIHTGEKPYTCEECGKAFNHSSHLATHKVIHTGEKPYQCEECGKAFNQSSHL
                     TRHKRIHTGEKPYQCEKCGKAFNQSSNLTGHKKIHTGEKLYKPKRCNSDFENTSKFSK
                     HKRNYAGEKS"
ORIGIN
```

```
  Query Match             8.0%;  Score 99.4;  DB 9;  Length 2132;
  Best Local Similarity   50.6%;  Pred. No. 1e-06;
  Matches  297;  Conservative    0;  Mismatches  281;  Indels    9;  Gaps    2;
```



RESULT 13
AC099500/c
LOCUS AC099500 138627 bp DNA linear PRI 29-MAY-2002
DEFINITION Homo sapiens chromosome 19 clone RP11-209J6, complete sequence.
ACCESSION AC099500
VERSION AC099500.2 GI:21240690
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 138627)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 138627)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 138627)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On May 29, 2002 this sequence version replaced gi:16930916.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www-shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;
Estimated Total Number of Errors is 0.1.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 153.8kb). It is clipped at the overlap with AC008981.
The number of bases overlapped is 27192.
FEATURES Location/Qualifiers
source 1. .138627
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-209J6"
ORIGIN

Query Match 8.0%; Score 99.4; DB 9; Length 138627;
Best Local Similarity 50.6%; Pred. No. 3.8e-07;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;



RESULT 14
AC138469/c
LOCUS AC138469 169500 bp DNA linear HTG 08-JAN-2003
DEFINITION Homo sapiens chromosome 19 clone RP11-189C24, *** SEQUENCING IN PROGRESS ***, 5 ordered pieces.
ACCESSION AC138469
VERSION AC138469.1 GI:27544967
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169500)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169500)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    75796: contig of 75796 bp in length
            *    75797    75896: gap of unknown length
            *    75897   138071: contig of 62175 bp in length
            *   138072   138171: gap of unknown length
            *   138172   157899: contig of 19728 bp in length
            *   157900   157999: gap of unknown length
            *   158000   166058: contig of 8059 bp in length
            *   166059   166158: gap of unknown length
            *   166159   169500: contig of 3342 bp in length.
            -------Genome Center
                        Center: Joint Genome Institute
                        Center Code: JGI
                        Web site: http://www.jgi.doe.gov
                        -------
                        Project Information
                        Center Project Name: 479013
                        Center clone name: RP11-189C24
                        -------
                        Summary Statistics
                        Consensus quality: 165586 bases at least Q40
                        Consensus quality: 165891 bases at least Q30
                        Consensus quality: 166156 bases at least Q20
                        Estimated insert size: 169095000; agarose-fp estimation
                        Estimated insert size: 166687; sum-of-contigs
            estimation
                        Quality coverage:  0 in Q20 bases; agarose-fp
            estimation
                        Quality coverage:  1 in Q20 bases; sum-of-contigs
            estimation.
FEATURES            Location/Qualifiers
     source         1. .169500
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="19"
                    /clone="RP11-189C24"
ORIGIN

  Query Match             8.0%;  Score 99.4;  DB 2;  Length 169500;
  Best Local Similarity   50.6%;  Pred. No. 3.6e-07;
  Matches  297;  Conservative    0;  Mismatches  281;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |          || ||  |  |  || |    |      |  | | |
Db      49667 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 49608

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
               | || |    | || || |||||  ||||  |||||  |    |||||  | |||  | |
Db      49607 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 49548

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |||||    | ||  || | |    ||   ||| || | ||||       || || |||
Db      49547 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 49494

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||  |      |||||  | |  |||||||  |||   ||| | | | |
Db      49493 GGCAAAGCCTTTAACCATTCCTCACACCTTGCTACACATAAGGTAATTCATACTGGAGAG 49434

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | ||| |||  ||  ||||||||||   ||||        ||||  | |  |||
Db      49433 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 49374

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |  |  ||||                |   || |||||| |||    |||| |||
Db      49373 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 49314

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              |  |       |  | |   ||||| | | | |  | |  ||| |||   | |
Db      49313 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 49254

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||||  |||||| |  | |||         |||||| ||  |  ||||||| |||
Db      49253 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 49194

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
              || |||||      ||    |      ||   |        ||||||| ||  ||
Db      49193 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 49134

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
              ||||| |      ||  ||     ||| | ||    | | | || |
Db      49133 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTATAAGT 49087


RESULT 15
AK122869
LOCUS       AK122869               2237 bp    mRNA    linear   PRI 09-SEP-2003
DEFINITION  Homo sapiens cDNA FLJ16502 fis, clone FEBRA2006664, moderately
            similar to Zinc finger protein 43.
ACCESSION   AK122869
VERSION     AK122869.1  GI:34528067
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M.,
            Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
            Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
            Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
            Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
            Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2237)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-2003) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES            Location/Qualifiers
     source         1. .2237
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="FEBRA2006664"
                    /tissue_type="brain"
                    /clone_lib="FEBRA2"
                    /dev_stage="fetal"
                    /note="cloning vector: pME18SFL3"
ORIGIN
```

Query Match 7.9%; Score 98.2; DB 9; Length 2237;
Best Local Similarity 52.0%; Pred. No. 1.6e-06;
Matches 273; Conservative 0; Mismatches 243; Indels 9; Gaps 2;

```
Qy        69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
             | || |    | || || |||||  ||||  |||||  |   |||||  | |||  | |
Db      1652 AGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTTAC 1711

Qy       129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
               |||||   | ||  || | |   ||   ||| || | ||||       || || ||||
Db      1712 TGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGTGG 1765

Qy       189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              ||||| |||  |       |||||  | |  |||||||  |||   ||| | | | | ||
Db      1766 CAAAGCCTTTAACCATTCCTCACACCTTGCTACACATAAGGTAATTCATACTGGAGAGAA 1825

Qy       249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
             ||| | ||| |||  ||  ||||||||||   ||||          |||| | |  |||||
Db      1826 ACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGACA 1885

Qy       309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
             | | |  |  |||||          |   || |||||| |||    |||| ||| |
Db      1886 TAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTTAA 1945

Qy       369 TAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
               |      |    | |  |||||| | | | |   |  |  ||| |||   | |
Db      1946 CCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAACC 2005

Qy       426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
             ||||  |||||| |  | |||         |||||||  ||  |  ||||||||  |||  |
Db      2006 TAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAATTA 2065

Qy       486 TCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAAC 545
             |  |||||     ||    |      ||    |       |||||||  ||  || ||
Db      2066 TGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCACAC 2125

Qy       546 TTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
             ||| |      ||  ||     |||  | ||    | | | || |
Db      2126 TTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTATAAGT 2170
```

Search completed: May 9, 2004, 08:56:45
Job time : 5087.84 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 03:47:06 ; Search time 544.339 Seconds
(without alignments)
9646.138 Million cell updates/sec

Title: US-09-831-804-1_COPY_720_1955
Perfect score: 1236
Sequence: 1 atgagtgaaagtgacgaaac.....aaacatcagtgatttctcga 1236

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1236	100.0	2060	3	AAA15398	Aaa15398 DNA encod
2	1232.8	99.7	1239	6	ABZ32206	Abz32206 Candida a
3	99.4	8.0	2132	7	ADA53158	Ada53158 Human cod
4	96.8	7.8	1705	8	ACA98970	Aca98970 cDNA enco
5	96.6	7.8	2320	7	ACA56456	Aca56456 Human sig
6	95.2	7.7	4563	5	AAS91317	Aas91317 DNA encod
7	94.6	7.7	2597	7	AAL51569	Aal51569 Human nuc
8	92.4	7.5	3639	5	AAS64586	Aas64586 DNA encod
9	91.8	7.4	2110	9	ADB63624	Adb63624 Human cDN
10	91.8	7.4	2760	9	ADC56695	Adc56695 Human mac
11	91.8	7.4	3078	9	ADB62468	Adb62468 Human cDN
12	90.4	7.3	2026	7	AAD55863	Aad55863 Human nuc
13	90.4	7.3	2114	7	ADA53124	Ada53124 Human cod
14	90.2	7.3	2230	7	ADA52931	Ada52931 Human cod
15	89.2	7.2	2064	9	ADC30762	Adc30762 Human nov
16	89.2	7.2	2622	5	AAS68872	Aas68872 DNA encod
17	89.2	7.2	2729	4	AAH16178	Aah16178 Human cDN
18	89.2	7.2	3839	6	ABK83826	Abk83826 Human cDN
19	89	7.2	2298	4	AAH16608	Aah16608 Human cDN
20	89	7.2	2905	5	AAS66143	Aas66143 DNA encod
21	89	7.2	3020	5	AAS92560	Aas92560 DNA encod
22	89	7.2	3502	4	AAI57845	Aai57845 Human pol
23	88.6	7.2	2597	7	AAD55855	Aad55855 Human nuc
24	88.6	7.2	4227	8	ACA98938	Aca98938 cDNA enco
25	88.4	7.2	1952	9	ADC58104	Adc58104 Zinc fing
26	88.4	7.2	2509	6	AAD31103	Aad31103 Human tra
27	87.6	7.1	1757	6	ABQ81135	Abq81135 TRAF6-inh
28	87.2	7.1	831	6	ABQ55095	Abq55095 Human ova
29	87	7.0	976	5	AAS69188	Aas69188 DNA encod
30	86.8	7.0	2662	6	ABQ93353	Abq93353 Human cDN
31	86.8	7.0	6219	7	ACC46324	Acc46324 Human dit
32	86.8	7.0	6316	7	ACC46373	Acc46373 Human dit
33	86.2	7.0	1377	7	ABX34443	Abx34443 Human mdd
34	86.2	7.0	2558	4	AAS26690	Aas26690 Human gen
c 35	86.2	7.0	2558	4	AAS26691	Aas26691 Human gen
36	86.2	7.0	2558	7	ABX74039	Abx74039 Human nov
c 37	86.2	7.0	2558	7	ABX74040	Abx74040 Human nov
38	86.2	7.0	2607	5	AAS87125	Aas87125 DNA encod
39	86.2	7.0	2681	7	ABX34772	Abx34772 Human mdd
40	86	7.0	1890	9	ADC30336	Adc30336 Human nov
41	86	7.0	3309	7	ACC46347	Acc46347 Human dit
42	86	7.0	3400	9	ADB62883	Adb62883 Human cDN
43	85.6	6.9	1549	5	AAS67562	Aas67562 DNA encod
44	85.6	6.9	2476	7	ACC46407	Acc46407 Human dit
45	84.8	6.9	2239	7	ADA53516	Ada53516 Human cod

ALIGNMENTS

RESULT 1
AAA15398
ID AAA15398 standard; DNA; 2060 BP.
XX
AC AAA15398;
XX
DT 04-SEP-2000 (first entry)
XX
DE DNA encoding a transcription factor designated CATFIIIA.
XX
KW Transcription factor; CATFIIIA; DNA-binding protein;
KW ribosomal RNA 5S gene; fungal infection; ss.
XX
OS Candida albicans.
XX
FH Key Location/Qualifiers
FT CDS 720. .1958
FT /*tag= a
FT /transl_except= (pos: 1296. .1298, aa: Ser)
FT /transl_except= (pos: 1734. .1736, aa: Ser)
XX
PN WO200028037-A1.
XX
PD 18-MAY-2000.
XX
PF 09-NOV-1999; 99WO-FR002739.
XX
PR 10-NOV-1998; 98FR-00014147.
XX
PA (HMRI) HOECHST MARION ROUSSEL.
XX
PI Bordon-Pallier F, Camier S, Sentenac A;
XX
DR WPI; 2000-376549/32.
DR P-PSDB; AAY93316.
XX
PT New nucleic acid encoding Candida albicans transcription factor, useful
PT e.g. in screening for antimycotic agents and for immunization.
XX
PS Claim 4; Page 32-33; 45pp; French.
XX
CC The present sequence encodes a Candida albicans transcription factor,
CC designated CATFIIIA. The polypeptide is a DNA-binding protein, which is
CC involved in initiating transcription of the ribosomal RNA 5S gene. The
CC polynucleotide is used to screen for its specific inhibitors, potentially
CC useful as antimycotic agents, to raise an antibody response that is

CC protective against fungal infection and to raise antibodies. Such
CC antibodies, as well as the polypeptides and polynucleotides are used in
CC compositions for diagnosing and treating fungal infections, e.g. by
CC detecting polymorphisms and mutations
XX
SQ Sequence 2060 BP; 726 A; 315 C; 354 G; 665 T; 0 U; 0 Other;

Query Match 100.0%; Score 1236; DB 3; Length 2060;
Best Local Similarity 100.0%; Pred. No. 1e-233;
Matches 1236; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 779

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 839

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 899

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 959

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 1019

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 1079

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 1139

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 1199

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 1259

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1260 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 1319

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 1379

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1380 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 1439

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1440 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 1499

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1500 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 1559

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1560 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 1619

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1620 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 1679

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1739

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1799

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1859

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1919

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236
              ||||||||||||||||||||||||||||||||||||
Db       1920 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1955
```

RESULT 2
ABZ32206
ID ABZ32206 standard; DNA; 1239 BP.
XX
AC ABZ32206;
XX
DT 30-JAN-2003 (first entry)
XX
DE Candida albicans essential gene SEQ ID NO 6493.
XX
KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
XX
OS Candida albicans.
XX
PN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US049486.
XX
PR 29-DEC-2000; 2000US-0259128P.
PR 20-FEB-2001; 2001US-00792024.
PR 22-AUG-2001; 2001US-0314050P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX
DR WPI; 2002-566694/60.
DR P-PSDB; ABP73656.
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele of
PT a gene and placing other allele of the gene under conditional expression.
XX
PS Claim 37; SEQ ID NO 6493; 167pp + Sequence Listing; English.
XX
CC The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene

CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans gene used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office
XX
SQ Sequence 1239 BP; 481 A; 171 C; 215 G; 372 T; 0 U; 0 Other;

Query Match 99.7%; Score 1232.8; DB 6; Length 1239;
Best Local Similarity 99.8%; Pred. No. 3.9e-233;
Matches 1234; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTATTGGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 180

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236
              |||||||||||||||||||||||||||||||||||
Db       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGA 1236
```

RESULT 3
ADA53158
ID ADA53158 standard; cDNA; 2132 BP.
XX
AC ADA53158;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human coding sequence, SEQ ID 726.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS Homo sapiens.
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-395539/38.
DR P-PSDB; ADA54797.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 726; 205pp; English.

```
XX
CC   The present invention relates to novel human secretory or membrane
CC   proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC   ADA54071). The coding sequences are useful in the gene therapy of
CC   diseases caused by abnormalities of the proteins, e.g. cancer,
CC   inflammatory diseases, osteoporosis or neurological disease.
XX
SQ   Sequence 2132 BP; 833 A; 423 C; 347 G; 529 T; 0 U; 0 Other;

  Query Match             8.0%;  Score 99.4;  DB 7;  Length 2132;
  Best Local Similarity   50.6%;  Pred. No. 3.4e-10;
  Matches  297;  Conservative    0;  Mismatches  281;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |           || ||  |  |  || |   |      |   | | |
Db       1384 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 1443

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
               | ||| |    | || || ||||  ||||  |||||  |   |||||  | |||  | |
Db       1444 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 1503

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |||||   | ||  || | |   ||   ||| || | ||||      || || |||
Db       1504 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 1557

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  |||||| |||  |       |||||  | |  |||||||  |||   ||| | | | |
Db       1558 GGCAAAGCCTTTAACCACTCCTCACACCTTGCTACACATAAGGTAATTCATACTGGAGAG 1617

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              |||||  | |||  |||   ||  ||||||||||   ||||          |||| | |  |||
Db       1618 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 1677

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||  |  |  |  |||||           |     || ||||||  |||    |||| |||
Db       1678 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 1737

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              |  |    |   |  |  |||||  | | |  |   |  |  |||  |||    |  |
Db       1738 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 1797

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||||  ||||||  |  | |||        |||||||  ||  |  ||||||||  |||
Db       1798 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 1857

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
              ||  |||||       ||    |       ||  |              |||||| ||  ||
Db       1858 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 1917

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
              |||||  |      ||  ||      |||  | ||    | | | || |
Db       1918 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTATAAGT 1964


RESULT 4
ACA98970
ID   ACA98970 standard; cDNA; 1705 BP.
XX
AC   ACA98970;
XX
DT   25-JUL-2003  (first entry)
XX
DE   cDNA encoding human nucleic acid-associated protein (NAAP) #51.
XX
KW   Human; nucleic acid-associated protein; cytostatic; antiarteriosclerotic;
KW   anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;
KW   antiallergic; antiinflammatory; thyromimetic; gene therapy;
KW   cell proliferative disorder; cancer; atherosclerosis;
KW   neurological disorder; epilepsy; Huntington's disease; stroke;
KW   immune disorder; inflammatory disorder; AIDS; allergy;
KW   developmental disorder; Hypothyroidism; Cushing's syndrome; infection;
KW   protein-protein interaction; drug-target interaction;
KW   gene expression profile; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003023003-A2.
XX
PD   20-MAR-2003.
XX
PF   05-SEP-2002; 2002WO-US028540.
XX
PR   07-SEP-2001; 2001US-0317792P.
PR   07-SEP-2001; 2001US-0317912P.
PR   14-SEP-2001; 2001US-0322270P.
PR   21-SEP-2001; 2001US-0324040P.
PR   28-SEP-2001; 2001US-0326732P.
PR   19-OCT-2001; 2001US-0346716P.
PR   25-JAN-2002; 2002US-0351749P.
PR   22-FEB-2002; 2002US-0359498P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Tang YT, Jackson JL, Griffin JA, Elliott VS, Forsythe IJ;
PI   Becha SD, Richardson TW, Lee EA, Sprague WW, Emerling BM;
PI   Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li JX;
PI   Hafalia AJA, Sanjanwala B, Marquis JP, Gorvad AE, Lee SY, Ison CH;
PI   Baughn MR, Chawla NK, Nguyen DB, Swarnakar A, Zebarjadian Y, Shah P;
PI   Thornton M, Yao MG, Khan FA, Gandhi AR, Yang J, Kable AE;
PI   Burford N, Ramkumar J;
XX
DR   WPI; 2003-313243/30.
DR   P-PSDB; ABU96722.
XX
PT   New human nucleic acid associated proteins (NAAP), useful for diagnosing,
PT   treating and preventing diseases or conditions associated with the
PT   aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
PT   infections.
XX
PS   Claim 5; Page 340-341; 345pp; English.
XX
CC   The invention describes a novel human isolated nucleic acid-associated
CC   polypeptide (NAAP). The polypeptides and polynucleotides are useful in
CC   diagnosing, treating and preventing diseases or conditions associated
CC   with the decreased expression or overexpression of NAAP, such as cell
CC   proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC   epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC   allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
CC   disorders, or infections. These are also useful in assessing the effects
CC   of exogenous compounds on the expression of nucleic acid and amino acid
CC   sequences of NAAP. The NAAP or its fragments are useful in screening
CC   compounds for effectiveness as agonist or antagonist of the polypeptides,
CC   or in altering the expression of the target polynucleotide and compounds
CC   that specifically bind to or modulate the activity of the polypeptide.
CC   The microarray is useful in monitoring or measuring protein-protein
CC   interactions, drug-target interactions, and gene expression profiles.
CC   This sequence encodes a novel human nucleic acid-associated protein
CC   (NAAP)
XX
SQ   Sequence 1705 BP; 622 A; 305 C; 314 G; 464 T; 0 U; 0 Other;

  Query Match             7.8%;  Score 96.8;  DB 8;  Length 1705;
  Best Local Similarity   52.9%;  Pred. No. 1.1e-09;
  Matches  259;  Conservative    0;  Mismatches  222;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||  | ||||    |  ||     ||    ||||  |      |    |   |
Db        695 GAATGTCGCAAAACATTTTGCATGCTTTCACACCTAACTCAACATAAAAGAATCCAAACT 754

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
               |  | || | | || || |||||  | ||  | ||||| |||  |  | |||   |
Db        755 AGAGTGAATTTCTACAAATGTGAAGCATATGGAAGAGCCTTTAACTGGTCCTCAACCCTT 814

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |  |||||   ||||  || |  ||   ||| || |||||||  |       || ||||
Db        815 AATAAACATAAGAGAATTCATACTGGAGAAAAACCTTACAAATGTAAA------GAATGT 868
```

```
Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
Db        869 GGCAAAGCCTTTAACCAGACCTCACACCTTATTAGACATAAGAGAATTCATACTGAAGAG 928

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
Db        929 AAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTCATCGACCCTTACTACA 988

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
Db        989 CATAATATAATTCATACTGGGGAAATTCCCTACAAATGTGAGAAATGTGTTAGAGCTTTT 1048

Qy        367 TATAAACATCAATCTTTAAGACATCATATATTATCTGTT---CATGAAAAAACATTAACG 423
Db       1049 AACCAAGCCTCAAAGCTTACTGAACATAAGTTAATTCATACCGGAGAGAAATGCTACAAA 1108

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
Db       1109 TGTGAAGAATGTGGCAAAGTTTTTAACCGATCTTCAAACCTTACAGACCATAAAAAAATT 1168

Qy        484 CATCATGGTG 493
Db       1169 CATACAGGAG 1178
```

RESULT 5
ACA56456
ID   ACA56456 standard; cDNA; 2320 BP.
XX
AC   ACA56456;
XX
DT   06-JUN-2003 (first entry)
XX
DE   Human signalling pathway polynucleotide probe SEQ ID NO 1054.
XX
KW   Human; probe; ss; array element; Parkinson's disease;
KW   signalling pathway population; cancer; adenocarcinoma; leukaemia;
KW   immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.
XX
OS   Homo sapiens.
XX
PN   US6500938-B1.
XX
PD   31-DEC-2002.
XX
PF   30-JAN-1998;   98US-00016434.
XX
PR   30-JAN-1998;   98US-00016434.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Au-Young J, Seilhamer JJ;
XX
DR   WPI; 2003-352189/33.
XX
PT   Combination of polynucleotide probes, useful as array elements in a
PT   microarray for monitoring the expression of a number of target
PT   polynucleotides.
XX
PS   Claim 1; SEQ ID NO 1054; 65pp; English.
XX
CC   The invention relates to a combination which, comprises a number of
CC   polynucleotide probes comprising a sequence selected from one of the 1490
CC   sequences mentioned in the specification. The combination is useful as an
CC   array element in a microarray for monitoring the expression of a number
CC   of target polynucleotides. The microarray is particularly useful in the
CC   diagnosis and treatment of cancer and immunopathology and neuropathology.
CC   The microarray is useful in diagnostics and treatment regimens, drug
CC   discovery and development, toxicological and carcinogenicity studies,
CC   forensics and pharmacogenomics. The microarray is also useful for
CC   monitoring progression of diseases and for developing sophisticated
CC   profiles for the effects of currently available therapeutic drugs. The
CC   combination is also useful for purifying a subpopulation of mRNAs, cDNAs
CC   and genomic fragments and in research and diagnostic applications. The
CC   array can detect changes in expression in a large number of genes coding
CC   for different signaling pathway populations which can be used to diagnose
CC   various diseases including cancer e.g. adenocarcinoma and leukaemia,
CC   immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
CC   and Parkinson's disease. The present sequence represents a polynucleotide
CC   probe of the invention. Note: The sequence data for this patent did not
CC   form part of the printed specification but was obtained in electronic
CC   format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=06500938B1
XX
SQ   Sequence 2320 BP; 903 A; 423 C; 412 G; 582 T; 0 U; 0 Other;

```
  Query Match             7.8%;  Score 96.6;  DB 7;  Length 2320;
  Best Local Similarity   51.3%;  Pred. No. 1.2e-09;
  Matches  306;  Conservative    0;  Mismatches  279;  Indels   12;  Gaps    3;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
Db        964 AGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTTAC 1023

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
Db       1024 TACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGTGG 1077

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
Db       1078 CAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAGAA 1137

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
Db       1138 ACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACACA 1197

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
Db       1198 TGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTTAA 1257

Qy        369 TAAACA---TCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
Db       1258 TCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAATG 1317

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
Db       1318 TAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATTCA 1377

Qy        486 TCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAAC 545
Db       1378 TACTGGAGAGAAGCCTTACAAATCTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCAAA 1437

Qy        546 TTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATG 605
Db       1438 ACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAATG 1494

Qy        606 TGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
Db       1495 TGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGA 1551
```

RESULT 6
AAS91317
ID   AAS91317 standard; cDNA; 4563 BP.
XX
AC   AAS91317;
XX
DT   13-FEB-2002 (first entry)
XX
DE   DNA encoding novel human diagnostic protein #27121.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.

XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG27130.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 27121; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4563 BP; 1752 A; 814 C; 824 G; 1173 T; 0 U; 0 Other;

Query Match 7.7%; Score 95.2; DB 5; Length 4563;
Best Local Similarity 52.7%; Pred. No. 2.5e-09;
Matches 258; Conservative 0; Mismatches 223; Indels 9; Gaps 2;

```
Qy        7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
            ||| ||| | || ||          || ||      |  ||      |    ||   | | |
Db     1315 GAATGTGGCAAAGCCTTTAACCAGTCCTCACACCTTACTCAACATAAAACAATTCATACT 1374

Qy       67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              | | |  | || || |||||  ||||  ||||| | ||| |  |  |||     |
Db     1375 GGAGGGAAAACCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCATCTTTCAGCCCTT 1434

Qy      127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
             |||||   | ||  || | |   ||   ||| || ||||||      || || |||
Db     1435 ACTAAACATAAGATAATTCATACTGGGGAGAAACCATACAAATGT------GAAGAATGT 1488

Qy      187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
             |  ||||| |||      |   |||||  |    | |||||  | |   ||| | | | |
Db     1489 GGCAAAGCTTTTAGGCAATCCTCACACCTTACTAGACATAAAGCAATTCATACTGGAGAG 1548

Qy      247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
            |||||  | | | |||  ||  |||||  ||||  |||||  |    |  || | | ||||
Db     1549 AAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCATTTCTCAGACCTTAGAAGA 1608

Qy      307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
```

```
                ||| | || |  |||||                |  | || |||| | |||    |||| |||
Db     1609 CATAAGATAATTCATACTGGAAAGAAACCCTACAAATGTGAAGAATGTGGGAAAGCTTTT 1668

Qy      367 TATAAACATCAATCTTTAAGA---CATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              |    | |  | |||   |||||  ||| |    |  |    || ||| | |  |
Db     1669 AGCCAGTCCTCAACCCTTAGAAACCATCAGATAATTCATACTGGAGAGAAACCCTACAAA 1728

Qy      424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
            ||| || |||||  |||||  ||| |   |  | ||||||  | |   |||||  |||
Db     1729 TGTGAAGAATGTGGTAAAGCTTTTAAGTGGTCATCAAAACTTACTGTACATAAGGTAATT 1788

Qy      484 CATCATGGTG 493
            |||  ||| |
Db     1789 CATACTGGAG 1798
```

RESULT 7
AAL51569
ID AAL51569 standard; DNA; 2597 BP.
XX
AC AAL51569;
XX
DT 10-APR-2003 (first entry)
XX
DE Human nucleic acid-associated protein coding sequence - SEQ ID No 52.
XX
KW Human; gene; ds; nucleic acid-associated protein; NAAP; arteriosclerosis;
KW cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
KW cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;
KW mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
KW Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
KW Crohn's disease; transgenic animal; animal model.
XX
OS Homo sapiens.
XX
PN WO2003000864-A2.
XX
PD 03-JAN-2003.
XX
PF 20-JUN-2002; 2002WO-US021179.
XX
PR 22-JUN-2001; 2001US-0300518P.
PR 29-JUN-2001; 2001US-0301787P.
PR 29-JUN-2001; 2001US-0301792P.
PR 29-JUN-2001; 2001US-0301892P.
PR 29-JUN-2001; 2001US-0301893P.
PR 06-JUL-2001; 2001US-0303405P.
PR 06-JUL-2001; 2001US-0303442P.
PR 15-MAR-2002; 2002US-0364438P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;
PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IJ;
PI Ramkumar J, Griffin JA, Yang J, Sanjanwala MM, Baughn MR;
PI Borowsky ML, Yao MG, Walia NK, Bandman O, Lal PG, Becha SD, Lee SY;
PI Richardson TW, Elliott VS, Luo W, Tang YT, Zebarjadian Y, Lu Y;
XX
DR WPI; 2003-201420/19.
DR P-PSDB; AAO16419.
XX
PT New nucleic acid-associated proteins and polynucleotides, useful for
PT diagnosing, treating or preventing cell proliferative (e.g. cancer),
PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
PT disorders (e.g. AIDS).
XX
PS Claim 12; Page 292-293; 312pp; English.
XX
CC The invention comprises the amino acid and coding sequences of human
CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
CC the invention are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of NAAP, such as: cell proliferative
CC disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis

```
CC   or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia
CC   or mental retardation); neurological disorders (e.g. Alzheimer's disease,
CC   Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders
CC   (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of
CC   the invention are useful for creating transgenic animals to model human
CC   disease. The present DNA sequence encodes a human nucleic acid-associated
CC   protein of the invention
XX
SQ   Sequence 2597 BP; 947 A; 504 C; 505 G; 641 T; 0 U; 0 Other;

  Query Match             7.7%;  Score 94.6;  DB 7;  Length 2597;
  Best Local Similarity   55.1%;  Pred. No. 3e-09;
  Matches  231;  Conservative    0;  Mismatches  179;  Indels    9;  Gaps    2;

Qy         78 TTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTT 137

Db       1070 TTATAAATGTGAAGAATGTGGCAAAGCCTTTAATGTGTCCTCAACCCTTACTCAACATAA 1129

Qy        138 AAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATT 197

Db       1130 GAGAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAGTGTGGCAAAGCCTT 1183

Qy        198 TTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCA 257

Db       1184 TAACGTGTCCTCAACTCTTACTCAACATAAGAGAATTCATACTGGAGAAAAACCATACAA 1243

Qy        258 TTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCAC 317

Db       1244 ATGTGAAGAATGTGGCAAAGCCTTTAACACATCCTCACACCTCACCACACATAAAAGAAT 1303

Qy        318 CCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAA---ACA 374

Db       1304 TCATACCGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTTCTC 1363

Qy        375 TCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATG 434

Db       1364 ACAACTTACTACACATAAGATAATTCATACTGGAGAGAAACCCTACAAATGTAAAGAATG 1423

Qy        435 TAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTG 493

Db       1424 TGGCAAAGCTTTTAAGCGGTCCTCAAACCTTACTGAACATAGGATAATTCATACTGGAG 1482
```

```
RESULT 8
AAS64586
ID   AAS64586 standard; cDNA; 3639 BP.
XX
AC   AAS64586;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #390.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG00399.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 390; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3639 BP; 1403 A; 708 C; 622 G; 906 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 92.4;  DB 5;  Length 3639;
  Best Local Similarity   54.1%;  Pred. No. 8.6e-09;
  Matches  236;  Conservative    0;  Mismatches  191;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128

Db        738 AGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGC 797

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188

Db        798 TAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 851

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248

Db        852 CAAAGCTTTTAGCCGTTCTTCAACCCTTGCTAAACATAAGAGAATTCATACTGGAGAGAA 911

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308

Db        912 ACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAGCAATTCCTCAACCCTTGCTAATCA 971

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368

Db        972 TAAGATAACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAGCTTTTAA 1031

Qy        369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425

Db       1032 GCGACTCTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATG 1091

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485

Db       1092 TGAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 1151

Qy        486 TCATGGTGGATCTCCT 501

Db       1152 TACTGGAGAGAAACCT 1167
```

```
RESULT 9
ADB63624
ID   ADB63624 standard; cDNA; 2110 BP.
XX
```

```
AC   ADB63624;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human cDNA encoding clone THYMU20071120.
XX
KW   Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW   tissue regeneration; cell regeneration; membrane protein;
KW   signal transduction-related protein; transcription-related protein;
KW   osteoporosis; neurological disease; cancer; tumour.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             170. .1792
FT                   /*tag=  a
FT                   /product= "Clone THYMU20071120 protein"
XX
PN   EP1308459-A2.
XX
PD   07-MAY-2003.
XX
PF   28-MAR-2002; 2002EP-00007401.
XX
PR   05-NOV-2001; 2001JP-00379298.
PR   25-JAN-2002; 2002US-00350978.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-450961/43.
DR   P-PSDB; ADB65594.
XX
PT   New polynucleotides and polypeptides, useful for developing a diagnostic
PT   marker or medicines for regulation of their expression and activity, or
PT   as targets of gene therapy.
XX
PS   Claim 1; Page; 222pp; English.
XX
CC   The invention discloses a polynucleotide comprising a sequence selected
CC   from 1970 fully defined nucleotide sequences which encode novel
CC   polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC   or its partial peptide, an antibody binding to the polypeptide or peptide
CC   of the polynucleotide, immunologically assaying the polypeptide or
CC   peptide of the polynucleotide by contacting the polypeptide or peptide
CC   with the antibody of the encoded protein, and observing the binding
CC   between the two, a transformant carrying the polynucleotide in an
CC   expressible manner and an antisense polynucleotide. The oligonucleotide
CC   is useful as a primer for synthesising the polynucleotide, or as a probe
CC   for detecting the polynucleotide. The polynucleotides and encoded
CC   proteins are useful as pharmaceutical agents and many disease-related
CC   genes may be included in them, for developing a diagnostic marker or
CC   medicines for regulation of their expression and activity, or as targets
CC   of gene therapy. The genes are involved in tissue and/or cell
CC   regeneration. Membrane proteins, signal transduction-related proteins,
CC   transcription-related proteins, disease-related proteins and genes
CC   encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC   neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC   the activity or expression of the encoded protein to treat diseases. The
CC   sequence presented is a cDNA of the invention. Note: Some of the sequence
CC   data for this patent is not represented in the printed specification, but
CC   is based on sequence information supplied by the European Patent Office.
XX
SQ   Sequence 2110 BP; 784 A; 394 C; 386 G; 546 T; 0 U; 0 Other;

  Query Match             7.4%;  Score 91.8;  DB 9;  Length 2110;
  Best Local Similarity   51.7%;  Pred. No. 1e-08;
  Matches  314;  Conservative    0;  Mismatches  277;  Indels   16;  Gaps    4;

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
               | | || || || |     | || || | ||| |||| ||||| | ||| |   | |||
Db       1253 CATACCGAAGAGAAACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAAACACTCCTCA 1312

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
               |      |||   |||| |||| |    ||   ||| || ||||||       ||
Db       1313 GCCCTTACTACACATAAGAGAATTCACACTGGAGAGAAACCCTACAAATGT------GAA 1366

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
               || ||||  ||||| ||| | ||  ||| | |     ||||    |   ||| |
Db       1367 GAATGTGGCAAAGCTTTTAACCGATCCTCAAAACTTACTGAACATAAGAAACTTCATACT 1426

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
               | ||| ||||| | | | ||| || ||||| ||||  |||       | | |
Db       1427 GGAAAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTATCCAATCCTCAAAACTT 1486

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
               |    |||| |||  |  ||| |          |   |  | | |||| | |||    ||
Db       1487 ACTGAACATAAAAAAATTCATTCTGGAGAGATACCCTACAAGTGTGAAGAATGTGGCAAA 1546

Qy        361 GCATTTTATAAACATCAATCTTTA---AGACATCATATATTATCTGTTCATGAAAAAACA 417
               || ||| |  |       ||| |   | |||| | | | |  |  | || ||| |
Db       1547 GCTTTTAAACACTCCTCATCCCTTACTACACATAAAAGAATTCATACTGGGGAGAAACCC 1606

Qy        418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477
               |  | ||| || ||||   |||| ||| |  ||| | |||||| |  | |||||||
Db       1607 TACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCGATCCTCAAAACTTACTGAACATAAG 1666

Qy        478 TTAAAACATCATGGTG-GATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAA 536
               |||  |||  ||| | ||  |||   ||    | |||   |    ||| |  ||
Db       1667 ATAATTCATACTGGAGAGAAACCCTATAAATGTGAGAGATGTGACAAAGCTTTTAACCAA 1726

Qy        537 TTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATG 596
               | | ||||| |      || || |  | | | |  |||| |  |||    | |
Db       1727 TCTGCAAACCT------TACTAAACATAAAAAAATACATACTGGAGAGAAACTACAGAAC 1780

Qy        597 TCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCA 656
               |    | |||| ||| | || ||      |    | ||| |||  | |||    || |||
Db       1781 TGGAATGTGTGATAATGATTTTGACAACACCTCAAATTTTTCTAAATATAAAGGAAATCA 1840

Qy        657 TGATGAT 663
               |  || |
Db       1841 TACTGGT 1847


RESULT 10
ADC56695
ID   ADC56695 standard; cDNA; 2760 BP.
XX
AC   ADC56695;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human macroprotein-45-76 cDNA.
XX
KW   human; macroprotein-45.76; dementia; arrhythmia; asthma; diabetes; ss;
KW   gene.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             180. .1430
FT                   /*tag=  a
FT                   /product= "Human macroprotein-45.76"
XX
PN   CN1382725-A.
XX
PD   04-DEC-2002.
XX
PF   26-APR-2001; 2001CN-00112764.
XX
PR   26-APR-2001; 2001CN-00112764.
XX
```

PA (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.
XX
PI Mao Y, Xie Y;
XX
DR WPI; 2003-269486/27.
DR P-PSDB; ADC56696.
XX
PT New human macroprotein-45.76, encoding polynucleotide, antagonist and
PT recombinant production, useful for treating dementia, arrhythmia, asthma
PT and diabetes.
XX
PS Claim 6; SEQ ID NO 1; 34pp; Chinese.
XX
CC The invention relates to a novel human macroprotein-45.76, the encoding
CC polynucleotide, an antagonist and a method of recombinant production. The
CC protein of the invention may be useful for treating dementia, arrhythmia,
CC asthma and diabetes. The current sequence is that of the human
CC macroprotein-45.76 cDNA of the invention.
XX
SQ Sequence 2760 BP; 1016 A; 463 C; 499 G; 782 T; 0 U; 0 Other;

Query Match 7.4%; Score 91.8; DB 9; Length 2760;
Best Local Similarity 52.1%; Pred. No. 1.1e-08;
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;

Qy 11 GTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCCAAAA 70
Db 712 GTGGCAAATCATTTTGCATGCTTTCACAATTAACTCAACATAAGAAAATTCATACTAGAG 771

Qy 71 AGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGC 130
Db 772 AGTATTCTTACAAATGTGAAGAATGTGGTAAAGCCTTTAACTGGTCCTCAACCCTTACTA 831

Qy 131 AACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATA 190
Db 832 AACATAAGATAATTCATACTGGAGAAAAACCCTACAAATGT------GAAGAATGTGGCA 885

Qy 191 AAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAC 250
Db 886 AAGCTTTTAACCGGTCCTCAAATCTTACTAAACATAAAATAATTCATACTGGAGAGAAAC 945

Qy 251 CATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATG 310
Db 946 CCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCGGTCCTCAACCCTTACTAAACATA 1005

Qy 311 AAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATA 370
Db 1006 AAAGAATTCATACAGAAGAGAAACCCTACAAATGTGAAGAATGTGGCAAGGCCTTTAACC 1065

Qy 371 AACATCAATCTTT---AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTA 427
Db 1066 AGTTCTCGATTCTTAATAAACATAAGAGAATTCATATGGAAGATAAACCCTACAAATGTG 1125

Qy 428 AACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATC 487
Db 1126 AAGAATGTGGCAAAGCCTTTAGAGTATTCTCAATTCTTAAAAAACATAAGATAATCCATA 1185

Qy 488 ATGGTGGATCTCC 500
Db 1186 CTGGGGAAAAACC 1198

RESULT 11
ADB62468
ID ADB62468 standard; cDNA; 3078 BP.
XX
AC ADB62468;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human cDNA encoding clone FEBRA20063720.
XX
KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW tissue regeneration; cell regeneration; membrane protein;
KW signal transduction-related protein; transcription-related protein;
KW osteoporosis; neurological disease; cancer; tumour.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 118. .1878
FT /*tag= a
FT /product= "Clone FEBRA20063720 protein"
XX
PN EP1308459-A2.
XX
PD 07-MAY-2003.
XX
PF 28-MAR-2002; 2002EP-00007401.
XX
PR 05-NOV-2001; 2001JP-00379298.
PR 25-JAN-2002; 2002US-00350978.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-450961/43.
DR P-PSDB; ADB64438.
XX
PT New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
PS Claim 1; Page; 222pp; English.
XX
CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.
XX
SQ Sequence 3078 BP; 1159 A; 516 C; 538 G; 865 T; 0 U; 0 Other;

Query Match 7.4%; Score 91.8; DB 9; Length 3078;
Best Local Similarity 52.1%; Pred. No. 1.1e-08;
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;

Qy 11 GTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCCAAAA 70
Db 650 GTGGCAAATCATTTTGCATGCTTTCACAATTAACTCAACATAAGAAAATTCATACTAGAG 709

Qy 71 AGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGC 130
Db 710 AGTATTCTTACAAATGTGAAGAATGTGGTAAAGCCTTTAACTGGTCCTCAACCCTTACTA 769

```
Qy        131 AACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATA 190
              |||||   | ||  || | |   ||   ||| || ||||||      || || ||||  |
Db        770 AACATAAGATAATTCATACTGGAGAAAAACCCTACAAATGT------GAAGAATGTGGCA 823

Qy        191 AAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAC 250
              ||||  |||   | |     ||| ||  |    | |||||   ||   ||| | | | ||||
Db        824 AAGCTTTTAACCGGTCCTCAAATCTTACTAAACATAAAATAATTCATACTGGAGAGAAAC 883

Qy        251 CATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATG 310
              | | | | |||  ||  ||||| ||||   ||||          |  | | | | ||||
Db        884 CCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCGGTCCTCAACCCTTACTAAACATA 943

Qy        311 AAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATA 370
              |||  |  ||||||  |            |  | || |||| | |||     | || |||| |
Db        944 AAAGAATTCATACAGAAGAGAAACCCTACAAATGTGAAGAATGTGGCAAGGCCTTTAACC 1003

Qy        371 AACATCAATCTTT---AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTA 427
              |     | |    | |||| | | | |   | |  | || ||| | |   |  |||
Db       1004 AGTTCTCGATTCTTAATAAACATAAGAGAATTCATATGGAAGATAAACCCTACAAATGTG 1063

Qy        428 AACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATC 487
              || |||||   ||||  || |    |  ||||   |   | |||||||  |||  |||
Db       1064 AAGAATGTGGCAAAGCCTTTAGAGTATTCTCAATTCTTAAAAAACATAAGATAATCCATA 1123

Qy        488 ATGGTGGATCTCC 500
               |||  | |   ||
Db       1124 CTGGGGAAAAACC 1136
```

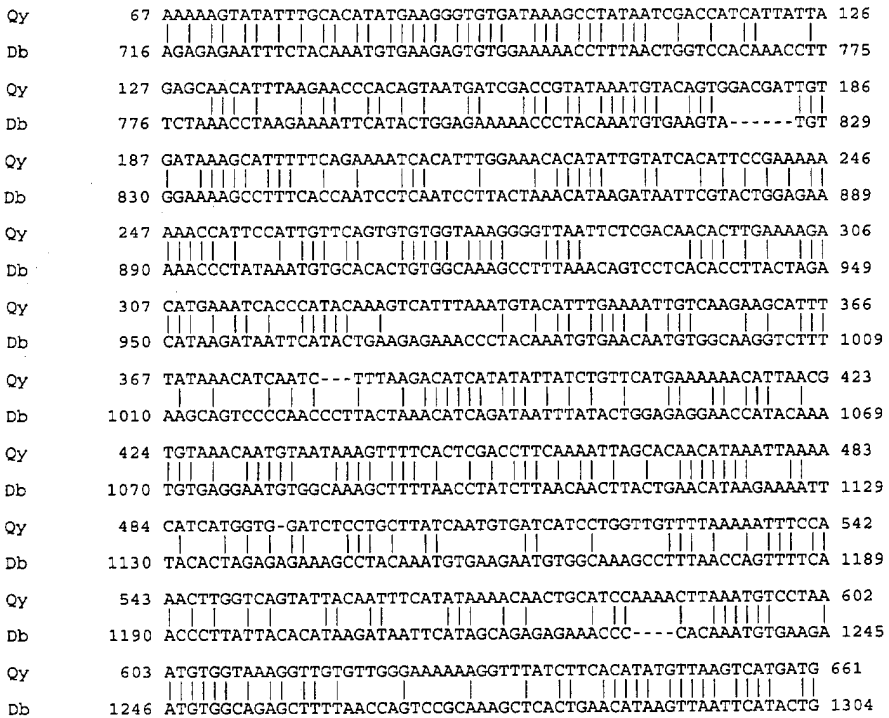
RESULT 12
AAD55863
ID AAD55863 standard; cDNA; 2026 BP.
XX
AC AAD55863;
XX
DT 07-AUG-2003 (first entry)
XX
DE Human nucleic acid associated protein (NAAP)-34 cDNA.
XX
KW Human; nucleic acid associated protein; NAAP; stroke; AIDS; nootropic;
KW cancer; atherosclerosis; neurological; epilepsy; Huntington's disease;
KW developmental disorder; antiinflammatory; neuroprotective; thyromimetic;
KW Cushing's syndrome; infection; gene therapy; cytostatic; anticonvulsant;
KW cerebroprotective; hypothyroidism; cell proliferative disorder; allergy;
KW gene; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 126. .1973
FT /*tag= a
FT /product= "Human NAAP protein"
XX
PN WO2003006618-A2.
XX
PD 23-JAN-2003.
XX
PF 10-JUL-2002; 2002WO-US021971.
XX
PR 12-JUL-2001; 2001US-0305089P.
PR 12-JUL-2001; 2001US-0305104P.
PR 13-JUL-2001; 2001US-0305325P.
PR 13-JUL-2001; 2001US-0305390P.
PR 19-JUL-2001; 2001US-0306960P.
PR 20-JUL-2001; 2001US-0306694P.
PR 27-JUL-2001; 2001US-0308170P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Swarnakar A, Richardson TW, Warren BA, Griffin JA, Tang YT;
PI Yue H, Baughn MR, Emerling BM, Lal PG, Lu DAM, Forsythe IJ;
PI Ramkumar J, Li JX, Becha SD, Duggan BM, Sanjanwala MM, Lee EA;
PI Burford N, Elliott VS, Ison CH, Ding L, Borowsky ML, Yao MG;
PI Barroso I, Tran B, Walia NK, Hafalia AJA, Nguyen DB, Lu Y;
PI Arvizu CS;
XX
DR WPI; 2003-221732/21.
DR P-PSDB; AAE37047.
XX
PT New human nucleic acid associated proteins (NAAP), useful for diagnosing,
PT treating and preventing diseases or conditions associated with the
PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
PT infections.
XX
PS Claim 5; Page 259-260; 260pp; English.
XX
CC The invention relates to human nucleic acid associated proteins (NAAP)
CC and their corresponding nucleic acid sequences. The invention is useful
CC in diagnosing, treating and preventing diseases or conditions associated
CC with the decreased expression or overexpression of NAAP, such as cell
CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
CC disorders, or infections. It is also useful in assessing the effects of
CC exogenous compounds on the expression of nucleic acid and amino acid
CC sequences of NAAP. The NAAP or its fragments are useful in screening
CC compounds for which acts as their agonist or antagonist. The microarray
CC is useful in monitoring or measuring protein-protein interactions, drug-
CC target interactions, and gene expression profiles. NAAP DNA is used in
CC gene therapy. The present sequence is human NAAP cDNA
XX
SQ Sequence 2026 BP; 699 A; 422 C; 396 G; 509 T; 0 U; 0 Other;

Query Match 7.3%; Score 90.4; DB 7; Length 2026;
Best Local Similarity 51.8%; Pred. No. 2e-08;
Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||  |    || ||      |  || |   |     |  | | |
Db       1002 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 1061

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              | || |    | || || |||||  |||| ||||||  | ||| |  || |||     |
Db       1062 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 1121

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |||         |  ||   ||| ||   ||| || ||||||      || || |||
Db       1122 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 1175

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || |||||  |||  | ||   ||| |  |    | |||||   ||   ||||| | | |
Db       1176 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 1235

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              |||  |  |  | |||  |   ||||| |||||  ||||  |        |  | | | || |
Db       1236 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 1295

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||  ||  |  |||||          |  | || |||| | |||     |||| |||
Db       1296 CATAGAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 1355

Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              ||      |  ||   |  | |||| | ||  |   |  |  || ||| | |  |
Db       1356 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 1415

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   |||| || |  |   | |||||| | |  ||||||  |||
Db       1416 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 1475

Qy        484 CATCATGGTGGATCTCCT 501
              |||  ||| |     |||
Db       1476 CATACTGGAGAGAAACCT 1493
```

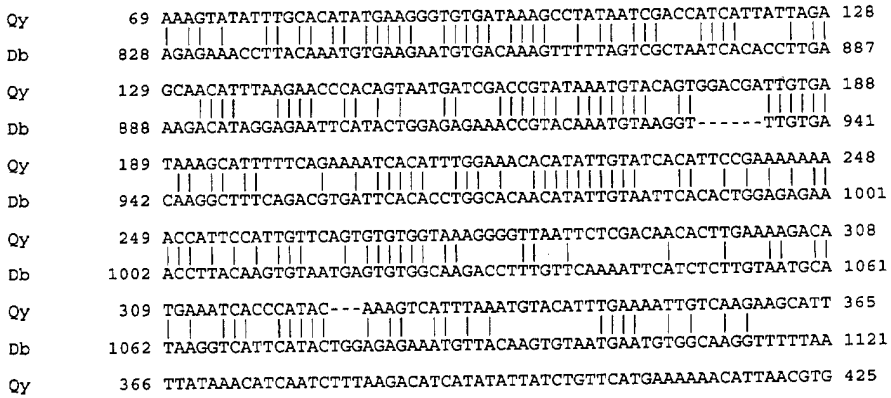
RESULT 13

```
ADA53124
ID   ADA53124 standard; cDNA; 2114 BP.
XX
AC   ADA53124;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human coding sequence, SEQ ID 692.
XX
KW   Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW   Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW   inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS   Homo sapiens.
XX
PN   EP1293569-A2.
XX
PD   19-MAR-2003.
XX
PF   21-MAR-2002; 2002EP-00006586.
XX
PR   14-SEP-2001; 2001JP-00328381.
PR   24-JAN-2002; 2002US-0350435P.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-395539/38.
DR   P-PSDB; ADA54763.
XX
PT   New polynucleotides encoding full-length polypeptides, e.g. secretory
PT   and/or membrane proteins, useful for developing medicines for diseases in
PT   which the gene is involved, or as target molecules for gene therapy.
XX
PS   Claim 1; SEQ ID NO 692; 205pp; English.
XX
CC   The present invention relates to novel human secretory or membrane
CC   proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC   ADA54071). The coding sequences are useful in the gene therapy of
CC   diseases caused by abnormalities of the proteins, e.g. cancer,
CC   inflammatory diseases, osteoporosis or neurological disease.
XX
SQ   Sequence 2114 BP; 717 A; 439 C; 434 G; 524 T; 0 U; 0 Other;

  Query Match             7.3%;  Score 90.4;  DB 7;  Length 2114;
  Best Local Similarity   51.8%;  Pred. No. 2e-08;
  Matches  258;  Conservative    0;  Mismatches  231;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||   |     || ||     | || |   |   |   | | |
Db       1024 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 1083

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | ||| |    | || || ||||| |||| |||||| | ||| |  || |||    |
Db       1084 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 1143

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |||       |  ||   ||| ||   ||| || ||||||      || || |||
Db       1144 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 1197

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || ||||| |||  | ||  ||| |  |    | |||||   ||   ||||| | | |
Db       1198 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 1257

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||| | | | | |||  |   ||||| |||||  |||| |     |  | | | || |
Db       1258 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 1317

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||  ||  |  |||||              |  | || |||| | |||    |||| |||
Db       1318 CATAGAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 1377

Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
                ||    |   ||   |   | |||| | ||  |   |  |   || ||| | |  |
Db       1378 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 1437

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   ||||  ||| |  |  | ||||||  |   |  ||||||  |||
Db       1438 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 1497

Qy        484 CATCATGGTGGATCTCCT 501
              |||  |||  |     |||
Db       1498 CATACTGGAGAGAAACCT 1515


RESULT 14
ADA52931
ID   ADA52931 standard; cDNA; 2230 BP.
XX
AC   ADA52931;
XX
DT   20-NOV-2003  (first entry)
XX
DE   Human coding sequence, SEQ ID 499.
XX
KW   Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW   Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW   inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS   Homo sapiens.
XX
PN   EP1293569-A2.
XX
PD   19-MAR-2003.
XX
PF   21-MAR-2002; 2002EP-00006586.
XX
PR   14-SEP-2001; 2001JP-00328381.
PR   24-JAN-2002; 2002US-0350435P.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-395539/38.
DR   P-PSDB; ADA54570.
XX
PT   New polynucleotides encoding full-length polypeptides, e.g. secretory
PT   and/or membrane proteins, useful for developing medicines for diseases in
PT   which the gene is involved, or as target molecules for gene therapy.
XX
PS   Claim 1; SEQ ID NO 499; 205pp; English.
XX
CC   The present invention relates to novel human secretory or membrane
CC   proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC   ADA54071). The coding sequences are useful in the gene therapy of
CC   diseases caused by abnormalities of the proteins, e.g. cancer,
CC   inflammatory diseases, osteoporosis or neurological disease.
XX
SQ   Sequence 2230 BP; 831 A; 434 C; 400 G; 565 T; 0 U; 0 Other;

  Query Match             7.3%;  Score 90.2;  DB 7;  Length 2230;
  Best Local Similarity   50.4%;  Pred. No. 2.2e-08;
  Matches  332;  Conservative    0;  Mismatches  313;  Indels   14;  Gaps    4;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||||| || |   |  ||   ||     || ||     |    |   | | |
Db        656 GAATGTGACAAATCACTTTGCATGCTTTCACGCCTAACTCAACATAAAAAAATTCATACT 715
```



RESULT 15
ADC30762
ID ADC30762 standard; cDNA; 2064 BP.
XX
AC ADC30762;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human novel cDNA sequence, SEQ ID NO:844.
XX
KW Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW gene therapy; chromosome 19q12; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003029271-A2.
XX
PD 10-APR-2003.
XX
PF 24-SEP-2002; 2002WO-US030474.
XX
PR 24-SEP-2001; 2001US-0324631P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Haley-Vicente D, Drmanac RT;
XX
DR WPI; 2003-371981/35.
DR P-PSDB; ADC31733.
XX
PT New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT cancer.
XX
PS Claim 1; SEQ ID NO 844; 1185pp; English.
XX
CC The invention relates to 971 novel human cDNA sequences (ADC29919-
CC ADC30889)and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of peventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC -ADC33394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a specifically
CC claimed human cDNA sequence of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2064 BP; 703 A; 375 C; 449 G; 537 T; 0 U; 0 Other;

Query Match 7.2%; Score 89.2; DB 9; Length 2064;
Best Local Similarity 53.7%; Pred. No. 3.4e-08;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;



```
Db       1122 TCACAAATCAAACCTTGCATGTCATCATAGACTTCATACTGGAGAGAAACCTTACAAGTG 1181

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485

Db       1182 TAATGAATGTGGCAAGGTTTTTAATTGAAAATCAAACCTTGAACATCATCATAGAGTTCA 1241

Qy        486 TCATGGTGGATCTCCT 501

Db       1242 TATCGGAGAAAAACCT 1257
```

Search completed: May 9, 2004, 06:07:03
Job time : 557.339 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 05:31:45 ; Search time 3610.62 Seconds
(without alignments)
10222.531 Million cell updates/sec

Title: US-09-831-804-1_COPY_720_1955
Perfect score: 1236
Sequence: 1 atgagtgaaagtgacgaaac.....aaacatcagtgatttctcga 1236

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	182.2	14.7	628	28	BZ298000	BZ298000 CG3957.f1
c	2	167.6	13.6	959	29	CNS06MMP	AL405671 T7 end of
	3	98.2	7.9	1609	11	AK032220	AK032220 Mus muscu
	4	97.8	7.9	2647	11	BC047646	BC047646 Homo sapi
	5	97.6	7.9	745	9	AU123448	AU123448 AU123448
	6	92.4	7.5	2672	11	BC036394	BC036394 Homo sapi
	7	92	7.4	591	28	AQ005136	AQ005136 CIT-HSP-2
	8	91	7.4	593	28	B99387	B99387 CIT-HSP-228
	9	91	7.4	1163	13	BX414627	BX414627 BX414627
	10	90.4	7.3	2694	11	BC022527	BC022527 Homo sapi
	11	90.4	7.3	2698	11	BC037782	BC037782 Homo sapi
	12	90.2	7.3	2253	11	BC028252	BC028252 Mus muscu
	13	89.2	7.2	604	10	BE161630	BE161630 MR3-HT044
	14	89.2	7.2	3609	11	BC032590	BC032590 Homo sapi
	15	88.6	7.2	535	9	AL705393	AL705393 DKFZp686M
	16	88.6	7.2	2174	11	AK033001	AK033001 Mus muscu
	17	88.6	7.2	2330	11	BC037426	BC037426 Homo sapi
	18	88.6	7.2	2622	11	AK033958	AK033958 Mus muscu
	19	88.4	7.2	1084	13	BX456765	BX456765 BX456765
	20	88	7.1	895	13	BU195416	BU195416 AGENCOURT
	21	87.6	7.1	832	13	BX437291	BX437291 BX437291
	22	87.6	7.1	1201	13	BX355654	BX355654 BX355654
	23	87.4	7.1	1200	13	BX437758	BX437758 BX437758
	24	87.2	7.1	579	28	AQ347265	AQ347265 RPCI11-11
c	25	87.2	7.1	687	28	AQ389266	AQ389266 RPCI11-15
	26	87.2	7.1	907	13	BQ423752	BQ423752 AGENCOURT
	27	87	7.0	660	14	CD770140	CD770140 AGENCOURT
	28	87	7.0	842	13	BU507408	BU507408 AGENCOURT
c	29	86.8	7.0	521	28	AQ475626	AQ475626 CITBI-E1-
	30	86.8	7.0	705	28	AQ194282	AQ194282 RPCI11-61
c	31	86.8	7.0	860	28	AQ749175	AQ749175 HS_5575_A
	32	86.8	7.0	897	13	BQ431141	BQ431141 AGENCOURT
	33	86.8	7.0	1906	11	AK013043	AK013043 Mus muscu
	34	86.6	7.0	694	12	BG696505	BG696505 602658829
	35	86.4	7.0	874	13	BU171453	BU171453 AGENCOURT
	36	86.4	7.0	2791	11	BC020045	BC020045 Homo sapi
c	37	86.2	7.0	523	14	CA874049	CA874049 K0929E01-
c	38	86	7.0	557	10	BF817669	BF817669 MR2-CI012
	39	86	7.0	667	14	CF135920	CF135920 UI-HF-BN0
c	40	86	7.0	685	29	AG149704	AG149704 Pan trogl
	41	86	7.0	801	14	CD657347	CD657347 AGENCOURT
	42	85.6	6.9	711	14	CB154257	CB154257 K-EST0212
	43	85.6	6.9	712	9	AI792344	AI792344 an30d02.y
	44	85.4	6.9	682	12	BG818171	BG818171 602779744
	45	85.4	6.9	2106	11	AK030776	AK030776 Mus muscu

ALIGNMENTS

RESULT 1
BZ298000/c
LOCUS BZ298000 628 bp DNA linear GSS 31-OCT-2002
DEFINITION CG3957.f1 Candida glabrata Random Genomic Library Candida glabrata genomic clone CG3957, genomic survey sequence.
ACCESSION BZ298000
VERSION BZ298000.1 GI:24440936
KEYWORDS GSS.
SOURCE Candida glabrata
ORGANISM Candida glabrata
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 628)
AUTHORS Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
TITLE Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast Candida glabrata
JOURNAL Genome Biol. 4 (2), R10 (2003)
MEDLINE 22508158
PUBMED 12620120
COMMENT Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie
Class: plasmid ends.

```
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Candida glabrata"
                     /mol_type="genomic DNA"
                     /strain="CBS 138"
                     /db_xref="taxon:5478"
                     /clone="CG3957"
                     /clone_lib="Candida glabrata Random Genomic Library"
ORIGIN

  Query Match             14.7%;  Score 182.2;  DB 28;  Length 628;
  Best Local Similarity   59.5%;  Pred. No. 3.2e-24;
  Matches  348;  Conservative    0;  Mismatches  228;  Indels    9;  Gaps    2;

Qy         84 ATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGAAC 143
              ||| ||    ||||| || |||| ||   | || || ||  |     ||||    |  ||
Db        625 ATACGATAATTGTGACAAGGCCTTTACAAGGCCGTCCTTGCTCACAGAACACCAGAATAC 566

Qy        144 CCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAG 203
                 ||  | |   | |    || | |     | | ||||| |  |||||   |
Db        565 ---AGTGCATCTTGGGAGGAAACCTTGGAAATGTAATCAATGTGAAAGTTCATTTACTAA 509

Qy        204 AAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAACCATTCCATTGTTC 263
              |||     || || || | |||  |      ||||  | ||  ||| ||| ||  |||||||
Db        508 AAAGATCCACTTAGAGAGACACTTATACACACACAGATGAAAGACCGTTTTATTGTTC 449

Qy        264 AGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATAC 323
              | | ||||| || ||| | | | || | |||||  ||||| |||||||  | || || ||
Db        448 ATTTTGTGGGAAGGGGCTAATTACTAGGCAACAACTGAAACGACATGAGGTAACTCACAC 389

Qy        324 AAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTT 383
               || || || || |||  || ||||  |||| | ||    || || || || | |    ||
Db        388 CAAATCGTTCAATTGTGAATATGAAGGTTGTAATGAGAGTTTCTACAAGCACCCTCAATT 329

Qy        384 AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGT 443
              |||    |||||| ||| | |||||  | |||    |||  ||| |  ||||| || |||
Db        328 AAGGGCTCATATTTTAGCAGTTCATTTACAAAGTCTAAAATGTCACGAATGCAACAAAAG 269

Qy        444 TTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCAT------GGTGGATC 497
              |||     ||||||  | | |   | | || | |   |||||| |||      | ||
Db        268 TTTTCAAAGACCTTACAGACTCAAAAATCACATAGCCAAACACCATAATCCAGATGTTGT 209

Qy        498 TCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATT 557
                |||||||||||||                || ||||| ||| | |||| |||| ||||| | ||
Db        208 CAATGCTTATCAATGTACCTTCAGTGTTTGTTCTAAGAGTTTCAAAACATGGTCTGCGTT 149

Qy        558 ACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTG 617
              | ||| ||| | ||| |    ||||| ||  | ||||| |||| |||  |||||    ||
Db        148 AAGATTACATGTTAAAAATGATCATCCGAAGTTAAAATGCCCTATATGCAGTAAACCCTG 89

Qy        618 TGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
              ||||||| || | ||||||  |   || |||  ||  || |||||
Db         88 TGTTGGGGAAGACGGTTTAAATATGCACATGAAAATCCACGATGA 44


RESULT 2
CNS06MMP/c
LOCUS       CNS06MMP                 959 bp    DNA     linear   GSS 17-JUN-2001
DEFINITION  T7 end of clone AU0AA005F10 of library AU0AA from strain CBS 3082
            of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION   AL405671
VERSION     AL405671.1  GI:12168715
KEYWORDS    GSS.
SOURCE      Saccharomyces kluyveri
  ORGANISM  Saccharomyces kluyveri
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 959)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 959)
  AUTHORS   Neuveglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
            Gaillardin,C. and Casaregola,S.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 9.
            Saccharomyces kluyveri
  JOURNAL   FEBS Lett. 487 (1), 56-60 (2000)
  MEDLINE   20584719
   PUBMED   11152884
REFERENCE   3  (bases 1 to 959)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .959
                     /organism="Saccharomyces kluyveri"
                     /mol_type="genomic DNA"
                     /strain="CBS 3082"
                     /db_xref="taxon:4934"
                     /clone="AU0AA005F10"
                     /clone_lib="AU0AA"
                     /note="end : T7"
     misc_feature    complement(<71. .>958)
                     /note="similar to Saccharomyces cerevisiae ORF YPR186c [
                     PZF1 ; TFIIIA (transcription initiation factor) ]"
                     /evidence=not_experimental
ORIGIN

  Query Match             13.6%;  Score 167.6;  DB 29;  Length 959;
  Best Local Similarity   57.2%;  Pred. No. 1.7e-21;
  Matches  324;  Conservative    1;  Mismatches  235;  Indels    6;  Gaps    1;

Qy        211 CATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAACCATTCCATTGTTCAGTGTGT 270
              || ||||||| ||||||  |  | |||||||| || |||||||| |||||:|    |||
Db        958 CACTTGGAAAGACATATGTTCACGCATTCCGATAATAAACCATTTCATTGTYCCTATTGT 899

Qy        271 GGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAAGTCA 330
              ||||||| |||| | |     || || || |||||||||| ||||| ||||| || ||
Db        898 GGTAAAGGTGTTACTACAAAGCAGCAGTTAAAAAGACATGAGATCACACATACCAAATCG 839

Qy        331 TTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTTAAGACAT 390
              || |||||  || |||    ||  | |||   ||||| || |||       ||||
Db        838 TTCAAATGTGAATATGAGGGCTGCAACGAATGTTTTTACAAGCATTCTCAACTAAGGTCA 779

Qy        391 CATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTTTCACT 450
              |||| | | || || ||     | | ||| || ||| | || ||   |||| | |||
Db        778 CATACACTTTCGGTACACTTGCAGAAATTGACCTGTGAGCATTGCGGTAAAATGTTCCAA 719

Qy        451 CGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCTCCTGCT------ 504
              | || |      |   | | |||     ||| |||||   ||   |  ||
Db        718 AGGCCATATCGTCTGCAAAACCATCTTGCAAAGCATCACAATGTGGATGTTGAAAACAAA 659

Qy        505 TATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTT 564
              |||||||| | | | ||     | | |||  |||| ||| | |   | ||
Db        658 TATCAATGTACTTACACGGCGTGCACAGAGACTTTTAAAACCTGGACTGCTCTTCAGCAG 599
```

```
Qy        565 CATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGG 624
              ||||| ||       ||||| ||  |  | |||     ||||| |||| |||||||||
Db        598 CATATCAAGGCGGACCATCCCAAGTTGCACTGTAACGTATGTGGCAAAGCTTGTGTTGGT 539

Qy        625 AAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATCAAAATATGG 684
                |   ||| ||      || ||| |  ||||| || ||    | || ||||  |||
Db        538 GAGTCTGGTCTACAGATGCACATGCGAGTCCATGACGAGGCTTTAGTAATAAAAAATTGG 479

Qy        685 ACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTTGAACATTAT 744
              |  |||     |||| |    |     ||||| |||||   |||  |  |    || ||
Db        478 AAATGTACCATTTGTCAGCAAGTCTCTTTTGCTAAGAAGGCTGATCTTTTATCTCACTAC 419

Qy        745 AATATCTTCCATGATGGTAATATCCC 770
              |       ||| | | | |||| ||
Db        418 ATGGAACATCATAAGGATGATATACC 393
```

RESULT 3
AK032220
LOCUS AK032220 1609 bp mRNA linear HTC 18-SEP-2003
DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length enriched library, clone:6430503O15 product:MSZF33 (FRAGMENT) homolog [Mus musculus], full insert sequence.
ACCESSION AK032220
VERSION AK032220.1 GI:26328050
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1609)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.

FEATURES Location/Qualifiers

```
source          1. .1609
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="FANTOM_DB:6430503O15"
                /db_xref="MGI:2395676"
                /db_xref="taxon:10090"
                /clone="6430503O15"
                /sex="male"
                /tissue_type="olfactory brain"
                /clone_lib="RIKEN full-length enriched mouse cDNA library"
                /dev_stage="adult"
CDS             330. .1223
                /note="unnamed protein product; MSZF33 (FRAGMENT) homolog
                [Mus musculus] (SPTR|O88251, evidence: FASTY, 87.1%ID,
                100%length, match=255)
                putative"
                /codon_start=1
                /protein_id="BAC27766.1"
                /db_xref="GI:26328051"
                /translation="MLETYRNLNAVGYNWEDSNIEEHCESSRRHGRHERNHTGEKPYE
                GIQYGEAFVHHSSLQMRKIIHTGEKRYKCNQCDKAYSRHSILQIHKRTHSGEKPYECN
                QCGKAFTQHSHLKIHMVTHTGEKPYKCNQCGKAFACHNKLQKHERTHTGEKPYKCDQC
                NKAFVYESYLQVHKKTHTGEKPYKCNECGKAFARHSHLKVHKITHTGEKPYKCNQCGK
                ALAYHSTLQVHQRTHTGEKPYECEQCGKAFANQSYFQVHKRIHTGEKPYKCDQCGKAF
                VGSSDLKRHERVHTGRETLQM"
polyA_signal    1594. .1599
                /note="putative"
polyA_site      1609
                /note="putative"
```

ORIGIN

Query Match 7.9%; Score 98.2; DB 11; Length 1609;
Best Local Similarity 51.2%; Pred. No. 1.5e-08;
Matches 310; Conservative 0; Mismatches 283; Indels 12; Gaps 3;

```
Qy         60 ACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATC 119
              || | |   | || |     | || ||   |    ||||||||||||||    ||||
Db        521 ACATACTGGAGAGAAACGCTACAAATGCAATCAATGTGATAAAGCCTATTCACGACACAG 580

Qy        120 ATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGA 179
                | || |   ||||  |||||| || |||    ||   ||| ||| ||||||      |
Db        581 CATTCTACAAATACATAAAAGAACACATAGTGGAGAGAAACCCTATGAATGTA------A 634

Qy        180 CGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTC 239
              | |||| |||||| |||     | |    ||| |  ||| |||||| ||  ||||| |
Db        635 TCAATGTGGTAAAGCCTTTACACAACACAGTCATCTCAAAATACATATGGTTACACATAC 694
```

```
Qy        240 CGAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTT 299
Db        695 TGGAGAGAAACCTTACAAATGTAATCAATGTGGTAAAGCCTTTGCATGTCATAATAAACT 754

Qy        300 GAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGA 359
Db        755 TCAAAAACATGAAAGAACACATACTGGAGAGAAACCTTACAAGTGTGATCAATGTAATAA 814

Qy        360 AGCATTT---TATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAAC 416
Db        815 AGCCTTTGTATATGAAAGTTATTTACAAGTTCATAAAAAAACACATACTGGAGAGAAACC 874

Qy        417 ATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAA 476
Db        875 TTACAAATGTAATGAATGTGGTAAAGCCTTTGCACGACACAGTCATCTCAAAGTGCATAA 934

Qy        477 ATTAAAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAA 536
Db        935 AATAACACATACTGGAGAGAAAC---CTTACAAATGTAATCAATGTGGTAAAGCCCTTGC 991

Qy        537 TTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATG 596
Db        992 ATATCATAGTACACTCCAAGTACATCAAAGAACACATACTGGAGAGAAGCCCTATGAATG 1051

Qy        597 TCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCA 656
Db       1052 TGAGCAATGTGGCAAAGCCTTTGCAAATCAAAGTTATTTCCAAGTACATAAAAGAATACA 1111

Qy        657 TGATG 661
Db       1112 TACTG 1116
```

RESULT 4
BC047646

LOCUS	BC047646 2647 bp mRNA linear HTC 19-NOV-2003
DEFINITION	Homo sapiens zinc finger protein 85 (HPF4, HTF1), mRNA (cDNA clone IMAGE:4828290), with apparent retained intron.
ACCESSION	BC047646
VERSION	BC047646.1 GI:29126804
KEYWORDS	HTC.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 2647)
AUTHORS	Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE	Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE	22388257
PUBMED	12477932
REFERENCE	2 (bases 1 to 2647)
AUTHORS	Strausberg,R.
TITLE	Direct Submission
JOURNAL	Submitted (03-MAR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT	Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: Miklos Palkovits, M.D., Ph.D. cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN) cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305 Web site: http://www-shgc.stanford.edu Contact: (Dickson, Mark) mcd@paxil.stanford.edu Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers, R. M. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 90 Row: i Column: 7 This clone has the following problem: retained intron.

FEATURES Location/Qualifiers

source	1. .2647
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="IMAGE:4828290"
	/tissue_type="Testis"
	/clone_lib="NIH_MGC_97"
	/lab_host="DH10B"
	/note="Vector: pBluescript"

ORIGIN

Query Match 7.9%; Score 97.8; DB 11; Length 2647;
Best Local Similarity 50.1%; Pred. No. 1.7e-08;
Matches 330; Conservative 0; Mismatches 317; Indels 12; Gaps 3;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
Db       1235 GAATGTGGCAAAACTTTTAACCGATTCTCAACTCTTACTACCCATAAGATAATTCATACT 1294

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
Db       1295 GGAGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTT 1354

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
Db       1355 ACTACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGT 1408

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
Db       1409 GGCAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAG 1468

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
Db       1469 AAACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACA 1528

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
Db       1529 CATGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTT 1588

Qy        367 TATAA---ACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
Db       1589 AATCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAA 1648

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
Db       1649 TGTAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATT 1708

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
Db       1709 CATACTGGAGAGAAGCCTTACAAATGTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCA 1768
```

```
Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAA 603
              |     | | |    |   |   | | | ||  |   | |||||   |||
Db       1769 AAACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAA 1825

Qy        604 TGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
              ||||| |||| || |    |     | | |  || |||| |  |||||||   ||
Db       1826 TGTGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGA 1884
```

RESULT 5
AU123448

```
LOCUS       AU123448               745 bp    mRNA    linear   EST 01-AUG-2002
DEFINITION  AU123448 NT2RM2 Homo sapiens cDNA clone NT2RM2000319 5', mRNA
            sequence.
ACCESSION   AU123448
VERSION     AU123448.1  GI:10948164
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 745)
  AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
            Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
            Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T.,
            Suzuki,Y., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1. .745
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RM2000319"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /clone_lib="NT2RM2"
                     /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
                     precursor cells"
ORIGIN
```

Query Match 7.9%; Score 97.6; DB 9; Length 745;
Best Local Similarity 50.2%; Pred. No. 2.1e-08;
Matches 328; Conservative 0; Mismatches 314; Indels 12; Gaps 3;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | ||||          ||| ||      |  || |    |      |  | | |
Db         87 GAATGTGGCAAAACTTTTAACCGATCCTCAACTCTTACTACCCATAAGATAATTCATACT 146

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | || |   | || || | |||  |||| |||||| | ||| ||  | |||     |
Db        147 GGAGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGTTCTTCAACCCTT 206

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                  |||  || ||  || | |   ||   ||| || ||||||     || || |||
Db        207 ACTACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGT 260

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              | ||||| |||          ||| | |   |||||||   ||   ||| | | | |
Db        261 GGCAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAG 320

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | | | |||  |   ||||| ||||   ||||          |||| | |  | |
Db        321 AAACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACTACA 380

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||||| | |  |||||                |  | ||| |||||| |||  | |||| |||
Db        381 CATGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTT 440

Qy        367 TATAA---ACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              ||| |      ||   |  | |||| | ||| |   |  |  || ||| | |  |
Db        441 AATCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAA 500

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              |||||| ||||| |||||  ||| |  |  ||||||   |  |  |||||||  |||
Db        501 TGTAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATT 560

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
              |||  ||| |     |||      |      ||  |        ||||||  | | |  |
Db        561 CATACTGGAGAGAAGCCTTACAAATGTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCA 620

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAA 603
              |     | | |    |   |   | | | ||  |   | |||||   |||
Db        621 AAACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAA 677

Qy        604 TGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCAT 657
              ||||| |||| || |    |     | | |  || |||| |  |||||||
Db        678 TGTGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCAT 731
```

RESULT 6
BC036394

```
LOCUS       BC036394               2672 bp    mRNA    linear   HTC 19-NOV-2003
DEFINITION  Homo sapiens zinc finger protein 85 (HPF4, HTF1), mRNA (cDNA clone
            IMAGE:5259399), containing frame-shift errors.
ACCESSION   BC036394
VERSION     BC036394.1  GI:23025784
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2672)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2672)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-AUG-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
```

```
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 73 Row: a Column: 16
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1. .2672
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5259399"
                     /tissue_type="Brain, hippocampus"
                     /clone_lib="NIH_MGC_95"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
ORIGIN
```

Query Match 7.5%; Score 92.4; DB 11; Length 2672;
Best Local Similarity 54.1%; Pred. No. 1.7e-07;
Matches 236; Conservative 0; Mismatches 191; Indels 9; Gaps 2;

```
Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
Db       1745 AGAGAAGCCTTACAAATGTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCAAAACTTAC 1804

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
Db       1805 TGAACATAAGAAAATTCATACTGGAGAGAAACCCTATGAATGT------GAAAAATGTGG 1858

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
Db       1859 CAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGAAGAGAA 1918

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
Db       1919 ACCTTACAAATGTGAAGAATGTGGCAAAGGTTTTAAATGGCCCTCAACCCTTACTATCCA 1978

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
Db       1979 TAAGATAATTCATACTGGAGAGAAACCATACAAATGTGAAGAATGTGGCAAAGCTTTTAA 2038

Qy        369 TAAA---CATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
Db       2039 CCAATCCTCAAAACTTACCAAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATG 2098

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
Db       2099 TGAAGAATGTGGCAAAGCCTTTAACCAGTCCTCAAACCTTACTAAACATAAGAGAATTCA 2158

Qy        486 TCATGGTGGATCTCCT 501
Db       2159 TACTGGAGAAAAACCT 2174
```

```
RESULT 7
AQ005136
LOCUS       AQ005136                591 bp    DNA     linear   GSS 27-JUN-1998
DEFINITION  CIT-HSP-2290O17.TF CIT-HSP Homo sapiens genomic clone 2290O17,
            genomic survey sequence.
ACCESSION   AQ005136
VERSION     AQ005136.1  GI:3082581
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2290O17.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21;
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .591
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7150844"
                     /db_xref="taxon:9606"
                     /clone="2290O17"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN
```

Query Match 7.4%; Score 92; DB 28; Length 591;
Best Local Similarity 52.2%; Pred. No. 2.4e-07;
Matches 256; Conservative 0; Mismatches 225; Indels 9; Gaps 2;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
Db         69 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 128

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
Db        129 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 188

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
Db        189 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 242

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
Db        243 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 302

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
Db        303 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 362

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
Db        363 CATAAAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 422

Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
Db        423 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 482

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
Db        483 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 542
```

```
Qy        484 CATCATGGTG 493
              ||| ||| |
Db        543 CATACTGGAG 552


RESULT 8
B99387
LOCUS       B99387                   593 bp    DNA     linear   GSS 26-JUN-1998
DEFINITION  CIT-HSP-2281A10.TF CIT-HSP Homo sapiens genomic clone 2281A10,
            genomic survey sequence.
ACCESSION   B99387
VERSION     B99387.1  GI:3027197
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 593)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21;
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .593
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7147044"
                     /db_xref="taxon:9606"
                     /clone="2281A10"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

  Query Match             7.4%;  Score 91;  DB 28;  Length 593;
  Best Local Similarity   54.4%;  Pred. No. 3.7e-07;
  Matches  231;  Conservative    0;  Mismatches  185;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              |||| |    | || || |||||  ||||  ||||| | ||  |    | |||    |
Db         47 AAAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCAGTCCTCAACCCTTAG 106

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
              |||||   | ||  || | |   ||   ||| || ||||||      || || ||||
Db        107 AAAACATGAGATAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 160

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              ||||||  |||   |    |||||  |    | |||||  |||   ||| | ||| | ||
Db        161 TAAAGCTTTTAAGTGGTCCTCACACCTTACTAGACATAAAGTAATTCATACTGAAGAGAA 220

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              |||  |  | |||  ||  |||||  || |   ||||   |      |  | | |  | |||
Db        221 ACCCTACAAATGTGAAGAATGTGGCAAGGCTTTTAACCATTTCTCAGCCCTTAGGAAACA 280

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | ||  |  |||||               |  | || |||| | |||    |||| |||
```

```
Db        281 TAAGATAATTCATACTGGAAAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAG 340

Qy        369 TAAACATCAATCTTT---AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
                |  || |   || |||| | ||| |   |  |  || ||| | |  |  ||
Db        341 CCAGTCCTCAACTCTTAGAAAACATGAGATAATTCATACTGGAGAGAAACCCTACAAATG 400

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||  ||||| ||| |   |  | ||||||  |  |   ||||||  |||  ||
Db        401 TGAAGAATGTGGTAAAGCTTTTAAGTGGTCCTCAAAACTTACTGTACATAAGGTAATTCA 460

Qy        486 TCATG 490
              |  ||
Db        461 TACTG 465


RESULT 9
BX414627
LOCUS       BX414627                 1163 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX414627 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP001YD12
            5-PRIME, mRNA sequence.
ACCESSION   BX414627
VERSION     BX414627.1  GI:30649903
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1163)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 9703.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0CAP001DB06QP1&cluster=9703.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0CAP001DB06QP1.
FEATURES             Location/Qualifiers
     source          1. .1163
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP001YD12"
                     /tissue_type="THYMUS"
                     /clone_lib="Homo sapiens THYMUS"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             7.4%;  Score 91;  DB 13;  Length 1163;
  Best Local Similarity   50.1%;  Pred. No. 3.5e-07;
  Matches  289;  Conservative    8;  Mismatches  268;  Indels   12;  Gaps    3;

Qy         89 AAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGAACCCACA 148
              |||  |||| |||||| | ||| ||| | |||    |      |||  |:|::  || |
Db        121 AAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTTACTACSCATAGARGRRTTCATA 180

Qy        149 GTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAAT 208
              |    ||  ||| || ||||||      || || ||:|  |||||  |||        |
Db        181 CTGGAGAGAAACCTTACAAATGT------GAAGAATGKGGCAAAGCCTTTAAGCAGTCCT 234

Qy        209 CACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGTTCAGTGT 268
              || |   |      |||||||    ||   ||| | | | | |||||  | | | |||  |    |
Db        235 CAAACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCCTACAAATGTAAAAGAT 294
```

```
Qy        269 GTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAAGT 328
              |||| ||||   |||          |||| | |  | |||||| | |  |||||
Db        295 GTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACACATGAGGTAATTCATACTGGAG 354

Qy        329 CATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAA---ACATCAATCTTTAA 385
              |   | || |||||  |||  | |||| ||| || |      ||    |   |
Db        355 AGAAACCCTACAAATGTGAAAGATGTGGAAAAGCCTTTAATCATTTCTCACACCTTACTA 414

Qy        386 GACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTT 445
              ||||  | ||| |   |  |  || ||| | |  |  |||||| :|||| ||||| ||
Db        415 CACATAAGATAATTCATACTGGAGAGAAACCTTACAAATGTAAAGRATGTGGTAAAGCTT 474

Qy        446 TCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCTCCTGCTT 505
              | |  |  ||||||   |  |  |||||||  |||  |||  |||  |    |||
Db        475 TTAAACACTCTTCAACCCTTACTAAACATAAGATAATTCATACTGGAGAGAAGCCTTACA 534

Qy        506 ATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTC 565
              |      |:   |   |  ||||||  | | |  ||      | | |   |
Db        535 AATGTAAAGRATGTGAAAGAGCTTTTAACCAATCCTCAAAACTTACTGAACATAAGARAA 594

Qy        566 ATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGA 625
              |  |    |  | | ||  |   | |||||   |:|||||| |||| ||  |     |
Db        595 TTCATACTGGAGAGAAACCCTA---TGAATGTGAAARATGTGGCAAAGCTTTTAACCAGT 651

Qy        626 AAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
              |  | |  ||  |||||  |  :||||||   ||
Db        652 CCTCAAATCTTACTAGACATAAGAGRAGTCATACAGA 688
```

RESULT 10
BC022527

```
LOCUS       BC022527                2694 bp    mRNA    linear   HTC 19-NOV-2003
DEFINITION  Homo sapiens hypothetical protein LOC148206, mRNA (cDNA clone
            IMAGE:4797857), containing frame-shift errors.
ACCESSION   BC022527
VERSION     BC022527.1  GI:18490307
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2694)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2694)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
```

```
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 32 Row: f Column: 2
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, Similarity but not identity to protein
            This clone has the following problem: frame shifted.
FEATURES              Location/Qualifiers
     source           1. .2694
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4797857"
                      /tissue_type="Brain, hypothalamus"
                      /clone_lib="NIH_MGC_96"
                      /lab_host="DH10B"
                      /note="Vector: pBluescript"
ORIGIN
```

```
  Query Match             7.3%;  Score 90.4;  DB 11;  Length 2694;
  Best Local Similarity   51.8%;  Pred. No. 4.1e-07;
  Matches  258;  Conservative    0;  Mismatches  231;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||   |    || ||     | || |    |    |   | | |
Db       1037 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 1096

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              | || |    | || || |||||  |||| ||||||  | ||| |   || |||     |
Db       1097 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTAACCACCCTTCAGCCCTT 1156

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              ||||      | ||   ||| ||  ||| || ||||||      || || |||
Db       1157 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 1210

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || ||||| ||| | ||   ||| |  |    | |||||   ||   ||||| | | |
Db       1211 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 1270

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||| | | | | ||| |  ||||| ||||  |||| |       | | | | || |
Db       1271 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 1330

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||  ||  |  ||||              | | || |||| | |||    |||| |||
Db       1331 CATAGAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 1390

Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              ||     |   ||  |  | |||| | |||  |  |  |  || ||| | | |
Db       1391 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 1450

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || ||||  ||||  ||  |  |  | |||||| |  |  ||||||  |||
Db       1451 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 1510

Qy        484 CATCATGGTGGATCTCCT 501
              |||  |||  |    |||
Db       1511 CATACTGGAGAGAAACCT 1528
```

```
RESULT 11
BC037782
LOCUS       BC037782                2698 bp    mRNA    linear   HTC 19-NOV-2003
DEFINITION  Homo sapiens hypothetical protein LOC148206, mRNA (cDNA clone
            IMAGE:4797729), with apparent retained intron.
ACCESSION   BC037782
VERSION     BC037782.1  GI:23349076
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2698)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2698)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-SEP-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 70 Row: g Column: 16
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Similarity but not
            identity to protein
            This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source          1. .2698
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4797729"
                     /tissue_type="Brain, hypothalamus"
                     /clone_lib="NIH_MGC_96"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
ORIGIN

  Query Match             7.3%;  Score 90.4;  DB 11;  Length 2698;
  Best Local Similarity   51.8%;  Pred. No. 4.1e-07;
  Matches  258;  Conservative    0;  Mismatches  231;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||   |    ||  ||    |  || |   |   |  | | |
Db       1042 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 1101

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
               | || |    | || || ||||  |||| ||||||  | ||| |  || |||    |
Db       1102 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 1161

Qy         127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                ||||     |  ||   ||| ||   ||| || ||||||      || || |||
Db       1162 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 1215

Qy         187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
               || ||||| |||  | ||  ||| |  |   | |||||  ||  |||||| | | |
Db       1216 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 1275

Qy         247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
               ||| | | | | |||   |  ||||| ||||  |||| |       | | | | || |
Db       1276 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 1335

Qy         307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
               |||  ||  |  ||||                |  | || |||| | |||    |||| |||
Db       1336 CATAGAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 1395

Qy         367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
               ||     |   ||   | |||| | ||| |  |   |  || ||| | |  |
Db       1396 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 1455

Qy         424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
               ||| || ||||  ||||  || |  |  | ||||||  |  |  ||||||  |||
Db       1456 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 1515

Qy         484 CATCATGGTGGATCTCCT 501
               |||  |||  |    |||
Db       1516 CATACTGGAGAGAAACCT 1533


RESULT 12
BC028252
LOCUS       BC028252                2253 bp    mRNA    linear   HTC 20-SEP-2002
DEFINITION  Mus musculus, clone IMAGE:3674739, mRNA.
ACCESSION   BC028252
VERSION     BC028252.1  GI:20380246
KEYWORDS    HTC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2253)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
            cDNA Library Preparation: Life Technologies, Inc.
```

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 65 Row: a Column: 17
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13386417
This clone has the following problem: frame shifted.

FEATURES Location/Qualifiers
source 1. .2253
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:3674739"
/tissue_type="Mammary tumor metastatized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."
/clone_lib="NCI_CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 7.3%; Score 90.2; DB 11; Length 2253;
Best Local Similarity 51.9%; Pred. No. 4.6e-07;
Matches 256; Conservative 0; Mismatches 228; Indels 9; Gaps 2;

```
Qy          4 AGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGT 63
              | | || ||||  || ||   || | |     ||   |    ||  |    | |||| |
Db        823 AATCAATGTGATAAAGCCTTTTCAACACACAGTTCTCTTCAAACTCATAGAAGAACACAT 882

Qy         64 CCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTA 123
               ||  || | |    | || || | |    |||||||||||||| |   ||   |   |
Db        883 ACCGGAAGGAAACCCTTCAAATGTAATCAATGTGATAAAGCCTTTTCCCGTAAACATTGT 942

Qy        124 TTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGAT 183
              | |    ||| |||| |  || | |   ||   ||| |  |||||||      | |
Db        943 GTGCAAACCCATATAAGGATACATACTGGAGAAAAACCCTTCAAATGTA------ATCAA 996

Qy        184 TGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAA 243
              |||||||||| ||||    | |     | |  |||| ||||    | |||||  ||| |
Db        997 TGTGATAAAGCCTTTTCACAAGACAGTACTCTTCAAACTCATAGAAGAACACATACCGGA 1056

Qy        244 AAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAA 303
              | |||||| ||| | |||      ||||| |||||   ||  |  |  |||      |  ||
Db       1057 GAGAAACCCTTCAAATGTAATCAGTGTGATAAAGCCTTTTCTGTTAAACATAGTCTCCAA 1116

Qy        304 AGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCA 363
              | ||||   |   |||||   |  |     | | || | |  | ||| |  ||||
Db       1117 ACACATATAGGAATACATACTGGGGAAAAACCCTACAAATGTAATCAATGTGATAAAGCC 1176

Qy        364 TTTTATAAACA---TCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||   ||||   ||||     ||  |||   | |  | |  |  ||| |  | |
Db       1177 TTTTCACAACACTTTCAACTCCAAAATCATATAAGAACACATACTAGAGAGATGCCTTAC 1236

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              |  ||||| |||||| | ||||  ||    ||||  |  |||  |       ||||   |
Db       1237 AAATGTAATCAATGTGACAAAGCCTTTTTACGACATAGAAATCTTCAGATTCATAGTAGA 1296

Qy        481 AAACATCATGGTG 493
```

```
              ||||||  ||| |
Db       1297 AAACATACTGGAG 1309
```

RESULT 13
BE161630
LOCUS BE161630 604 bp mRNA linear EST 21-JUN-2000
DEFINITION MR3-HT0446-260300-201-f06 HT0446 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE161630
VERSION BE161630.1 GI:8624351
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 604)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR3-HT0446-260300-201-f06&t3=2000-03-26&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 583.

FEATURES Location/Qualifiers
source 1. .604
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0446"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 7.2%; Score 89.2; DB 10; Length 604;
Best Local Similarity 53.7%; Pred. No. 8.1e-07;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

```
Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | ||  |   || || || |||||  |||||  ||||  |  ||  |||    ||||  | ||
Db        102 AGAGAAACCTTACAAATGTGAAGAATGTGACAAAGTTTTTAGTCGCTAATCACACCTTGA 161

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                ||||   ||||  || | |   ||   |||||| |||||||  ||       ||||||
Db        162 AAGACATAGGAGAATTCATACTGGAGAGAAACCGTACAAATGTAAGGT------TTGTGA 215

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              || || ||      |  | |||||  ||| |  ||||||||||    ||  | | | | ||
Db        216 CAAGGCTTTCAGACGTGATTCACACCTGGCACAACATATTGTAATTCACACTGGAGAGAA 275
```

```
Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||   | |||||| ||     ||  |       |    |   ||  ||
Db        276 ACCTTACAAGTGTAATGAGTGTGGCAAGACCTTTGTTCAAAATTCATCTCTTGTAATGCA 335

Qy        309 TGAAATCACCCATAC---AAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATT 365
              | | |||  |||||   | |||  || | |          ||||  | |  | |
Db        336 TAAGGTCATTCATACTGGAGAGAAATGTTACAAGTGTAATGAATGTGGCAAGGTTTTTAA 395

Qy        366 TTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
              | | |||     |   |  | | |||||||| | |   |  |   || ||| | |  | |||
Db        396 TCACAAATCAAACCTTGCATGTCATCATAGACTTCATACTGGAGAGAAACCTTACAAGTG 455

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              |||  |||||   || |||||  | | ||   |||||  | | |||  |||  |  |   ||
Db        456 TAATGAATGTGGCAAGGTTTTTAATTGAAAATCAAACCTTGAACATCATCATAGAGTTCA 515

Qy        486 TCATGGTGGATCTCCT 501
              |   || | |   |||
Db        516 TATCGGAGAAAAACCT 531
```

RESULT 14
BC032590
LOCUS BC032590 3609 bp mRNA linear HTC 19-NOV-2003
DEFINITION Homo sapiens cDNA clone IMAGE:5502691, containing frame-shift errors.
ACCESSION BC032590
VERSION BC032590.1 GI:21619671
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3609)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 3609)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 69 Row: g Column: 20
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13430873
This clone has the following problem: frame shifted.
FEATURES Location/Qualifiers
source 1. .3609
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5502691"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NIH_MGC_67"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
ORIGIN

Query Match 7.2%; Score 89.2; DB 11; Length 3609;
Best Local Similarity 53.7%; Pred. No. 6.7e-07;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

```
Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |   || || || |||||  |||||  ||||  | ||  |||   ||||   | ||
Db        587 AGAGAAACCTTACAAATGTGAAGAATGTGACAAAGTTTTTAGTCGCTAATCACACCTTGA 646

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                ||||   ||||  || | |   ||   ||||||  |||||||  ||      ||||||
Db        647 AAGACATAGGAGAATTCATACTGGAGAGAAACCGTACAAATGTAAGGT------TTGTGA 700

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              || || ||      | | |||||  ||| |  ||||||||||   ||  | | | | ||
Db        701 CAAGGCTTTCAGACGTGATTCACACCTGGCACAACATATTGTAATTCACACTGGAGAGAA 760

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||   | |||||| ||     ||  |       |    |   ||  ||
Db        761 ACCTTACAAGTGTAATGAGTGTGGCAAGACCTTTGTTCAAAATTCATCTCTTGTAATGCA 820

Qy        309 TGAAATCACCCATAC---AAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATT 365
              | |  ||| |||||   | ||  || | |          ||||  | |  | |
Db        821 TAAGGTCATTCATACTGGAGAGAAATGTTACAAGTGTAATGAATGTGGCAAGGTTTTTAA 880

Qy        366 TTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
              | | |||     |   |  | | |||||||| | |   |  |   || ||| | |  | |||
Db        881 TCACAAATCAAACCTTGCATGTCATCATAGACTTCATACTGGAGAGAAACCTTACAAGTG 940

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              |||  |||||   || |||||  | | ||   |||||  | | |||  |||  |  |   ||
Db        941 TAATGAATGTGGCAAGGTTTTTAATTGAAAATCAAACCTTGAACATCATCATAGAGTTCA 1000

Qy        486 TCATGGTGGATCTCCT 501
              |   || | |   |||
Db       1001 TATCGGAGAAAAACCT 1016
```

RESULT 15
AL705393
LOCUS AL705393 535 bp mRNA linear EST 04-SEP-2003
DEFINITION DKFZp686M1835_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone DKFZp686M1835_5', mRNA sequence.
ACCESSION AL705393
VERSION AL705393.1 GI:19688748

```
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   Ottenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Weil,B. and
            Wiemann,S.
  TITLE     EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and
            Wiemann,S.)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: MIPS
            MIPS
            Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by MediGenomix (Martinsried/Germany) within the cDNA
            sequencing consortium of the German Genome Project. No s1 sequence
            available.
            This clone (DKFZp686M1835) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1. .535
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686M1835"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="686 (synonym: hlcc3)"
                     /note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB;
                     cDNA-collection"
ORIGIN

  Query Match             7.2%;  Score 88.6;  DB 9;  Length 535;
  Best Local Similarity   53.3%;  Pred. No. 1.1e-06;
  Matches  237;  Conservative    0;  Mismatches  199;  Indels    9;  Gaps    2;

Qy         60 ACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATC 119
              ||  |  |    ||||  |    ||  ||  ||  |  |||  |||||  ||   |  |  |       |   ||
Db         86 ACATACTGGAAAGAAATCTTTCAAATGTAAAGAATGTGAAAAGTCATTTTGCATGCTTTC 145

Qy        120 ATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGA 179
              |    ||||   ||||||   |||||   ||  |||    ||    |||  ||  |||||||  |
Db        146 ACACTTAGCTCAACATAAAAGAATTCATAGTGGAGAGAAACCCTACAAATGTAAA----- 200

Qy        180 CGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTC 239
               ||  ||||   |||||  |  |        |   |||  |   |    |||||||     |    |||  |
Db        201 -GAATGTGGGAAAGCCTATAATGAGACCTCAAACCTTTCTACACATAAAAGAATTCATAC 259

Qy        240 CGAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTT 299
               | |||  |||||  |  |  |  ||    ||  ||||||  ||||    ||||      |     ||||  |
Db        260 TGGAAAGAAACCCTACAAATGCGAAGAGTGTGGAAAAGCCTTTAACCGGCTCTCACACCT 319

Qy        300 GAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGA 359
               |   |  ||||  |  ||  |   |||||                |   |  ||  |||   |  |||     |
Db        320 TACTACACATAAGATAATTCATACTGGAAAGAAACCCTACAAATGTGAGGAGTGTGGCAA 379

Qy        360 AGCATTTTATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAAC 416
              |||  |||  |   ||   |     |    |    | ||||  |  |  |  |   |    ||  |||  |
Db        380 AGCTTTTAACCAATCTGCAAACCTTACTACACATAAGAGAATTCATACTGGAGAGAAACC 439

Qy        417 ATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAA 476
               |   |  |||  ||  |||||     |  ||  |||  |   |    |  ||||    |   |    ||||||
Db        440 CTACAAATGTGAAGAATGTGGCAGAGCTTTTAGCCAGTCCTCAACCCTTACTGCACATAA 499

Qy        477 ATTAAAACATCATGGTGGATCTCCT 501
                 |||   |||   |||  |     |||
Db        500 GATAATTCATGCTGGAGAGAAACCT 524
```

```
Search completed: May  9, 2004, 10:57:39
Job time : 3629.62 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 03:49:53 ; Search time 5088.16 Seconds
(without alignments)
10554.310 Million cell updates/sec

Title: US-09-831-804-2
Perfect score: 1239
Sequence: 1 atgagtgaaagtgacgaaac.....catcagtgatttctcgataa 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1239	100.0	1239	6	BD274350	BD274350 Candida a
2		1239	100.0	2060	6	BD274349	BD274349 Candida a
3		1235.8	99.7	1239	6	AX489193	AX489193 Sequence
4		217.4	17.5	1560	8	YSCNOFEAT	M90638 Saccharomyc
5		217.4	17.5	1739	8	YSCTFIIIA	M80611 Saccharomyc
6	c	217.4	17.5	37497	8	YSCP9677	U25841 Saccharomyc
7	c	199.2	16.1	2133	8	YSCRP026A	M33924 S.cerevisia
8		193	15.6	797	11	CNS06ING	AL400514 T7 end of
9	c	101	8.2	169063	9	AC138126	AC138126 Homo sapi
10	c	101	8.2	177299	9	AC073544	AC073544 Homo sapi
11		99.4	8.0	2132	6	AX714042	AX714042 Sequence
12		99.4	8.0	2132	9	AK056088	AK056088 Homo sapi
13	c	99.4	8.0	138627	9	AC099500	AC099500 Homo sapi
14	c	99.4	8.0	169500	2	AC138469	AC138469 Homo sapi
15		98.2	7.9	2237	9	AK122869	AK122869 Homo sapi
16		97.8	7.9	1203	9	D70831	D70831 Homo sapien
17	c	97.8	7.9	235532	9	AC008739	AC008739 Homo sapi
18	c	97.2	7.8	115995	9	AC011494	AC011494 Homo sapi
19		96.6	7.8	2320	6	AR270491	AR270491 Sequence
20		96.6	7.8	2320	9	HSU35376	U35376 Human repre
21	c	96.6	7.8	68304	10	BX000432	BX000432 Mouse DNA
22		95	7.7	981	9	HSHZF9	X78932 H.sapiens H
23	c	94.6	7.6	186233	9	AC092329	AC092329 Homo sapi
24		94.6	7.6	189317	2	AC024483	AC024483 Homo sapi
25		94.2	7.6	292390	2	AC105677	AC105677 Rattus no
26		93.6	7.6	1389	9	M27879	M27879 Homo sapien
27	c	93.6	7.6	155439	9	AC008626	AC008626 Homo sapi
28		93.4	7.5	199585	10	AC124426	AC124426 Mus muscu
29	c	92.4	7.5	41153	9	AC016628	AC016628 Homo sapi
30		92.4	7.5	158430	2	AC012431	AC012431 Homo sapi
31		92	7.4	2873	9	HUMHPLK	M55422 Human Kruep
32		92	7.4	2873	11	G28705	G28705 sWSS3976 Er
33		92	7.4	2873	11	HUMSWS1269	G18281 human chrom
34		92	7.4	156835	9	AC010620	AC010620 Homo sapi
35	c	92	7.4	191082	2	AC141066	AC141066 Homo sapi
36	c	92	7.4	203396	9	AC073210	AC073210 Homo sapi
37		91.8	7.4	1870	9	AY044432	AY044432 Homo sapi
38		91.8	7.4	2110	6	AX748253	AX748253 Sequence
39		91.8	7.4	2110	9	AK093669	AK093669 Homo sapi
40		91.8	7.4	2909	9	HSM808021	BX647875 Homo sapi
41		91.8	7.4	2976	9	BC036439	BC036439 Homo sapi
42		91.8	7.4	3078	6	AX747097	AX747097 Sequence
43		91.8	7.4	3078	9	AK091618	AK091618 Homo sapi
44		91.8	7.4	3123	9	BC040594	BC040594 Homo sapi
45		91.8	7.4	174994	9	AC022145	AC022145 Homo sapi

ALIGNMENTS

RESULT 1
BD274350
LOCUS BD274350 1239 bp DNA linear PAT 17-JUL-2003
DEFINITION Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA protein.
ACCESSION BD274350
VERSION BD274350.1 GI:33084118
KEYWORDS JP 2002531068-A/2.
SOURCE Candida albicans
ORGANISM Candida albicans
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 1239)
AUTHORS Pallier,F.B., Camier,S. and Sentenac,A.
TITLE Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA

```
  JOURNAL   Patent: JP 2002531068-A 2 24-SEP-2002;
            AVENTIS PHARMA SA
COMMENT     OS   Candida albicans
            PN   JP 2002531068-A/2
            PD   24-SEP-2002
            PF   09-NOV-1999 JP 2000581204
            PR   10-NOV-1998 FR     98/14147
            PI   FLORENCE BORDON PALLIER,SYLVIE CAMIER,ANDRE SENTENAC PC
            C12N15/09,C12N15/09,A61K39/00,A61K45/00,A61P31/10,C07K14/40, PC
            C07K16/14,
            PC   C12N1/19,C12N1/21,C12P21/02,C12Q1/02,G01N33/15,G01N33/50, PC
            G01N33/53,
            PC   G01N33/566,G01N33/569//(C12N15/09,C12R1:725),C12N15/00,C12N15/
            PC   00,
            PC   (C12N15/00,C12R1:725)
            CC   Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
            CC           protein
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(1236).
FEATURES              Location/Qualifiers
     source           1. .1239
                      /organism="Candida albicans"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:5476"
ORIGIN

  Query Match           100.0%;  Score 1239;  DB 6;  Length 1239;
  Best Local Similarity  100.0%;  Pred. No. 1.7e-190;
  Matches 1239;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239
              ||||||||||||||||||||||||||||||||||||||
Db       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239


RESULT 2
BD274349
LOCUS       BD274349               2060 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
            protein.
ACCESSION   BD274349
VERSION     BD274349.1  GI:33084117
KEYWORDS    JP 2002531068-A/1.
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1  (bases 1 to 2060)
  AUTHORS   Pallier,F.B., Camier,S. and Sentenac,A.
  TITLE     Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
  JOURNAL   Patent: JP 2002531068-A 1 24-SEP-2002;
            AVENTIS PHARMA SA
COMMENT     OS   Candida albicans
            PN   JP 2002531068-A/1
            PD   24-SEP-2002
            PF   09-NOV-1999 JP 2000581204
            PR   10-NOV-1998 FR     98/14147
            PI   FLORENCE BORDON PALLIER,SYLVIE CAMIER,ANDRE SENTENAC PC
            C12N15/09,C12N15/09,A61K39/00,A61K45/00,A61P31/10,C07K14/40, PC
            C07K16/14,
            PC   C12N1/19,C12N1/21,C12P21/02,C12Q1/02,G01N33/15,G01N33/50, PC
            G01N33/53,
            PC   G01N33/566,G01N33/569//(C12N15/09,C12R1:725),C12N15/00,C12N15/
            PC   00,
            PC   (C12N15/00,C12R1:725)
            CC   Candida albicans tfIIIAgene (CatfIIIA) and the coded CATFIIIA
```

```
            CC              protein
            FH   Key             Location/Qualifiers
            FT   source          1. .2060
            FT                   /organism='Candida albicans'.
FEATURES              Location/Qualifiers
     source           1. .2060
                      /organism="Candida albicans"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:5476"
ORIGIN

  Query Match            100.0%;  Score 1239;  DB 6;  Length 2060;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-190;
  Matches 1239;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 779

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 839

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 899

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 959

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 1019

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 1079

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 1139

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 1199

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 1259

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1260 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 1319

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 1379

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1380 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 1439

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1440 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 1499

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1500 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 1559

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1560 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 1619

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1620 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 1679

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1739

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1799

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1859

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1919

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239
              ||||||||||||||||||||||||||||||||||||||
Db       1920 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1958


RESULT 3
AX489193
LOCUS       AX489193               1239 bp    DNA     linear   PAT 16-AUG-2002
DEFINITION  Sequence 6493 from Patent WO02053728.
ACCESSION   AX489193
VERSION     AX489193.1  GI:22323205
KEYWORDS    .
SOURCE      Candida albicans
  ORGANISM  Candida albicans
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE   1
  AUTHORS   Roemer,T., Jiang,B., Boone,C., Bussey,H. and Ohlsen,K.L.
  TITLE     Gene disruption methodologies for drug target discovery
  JOURNAL   Patent: WO 02053728-A 6493 11-JUL-2002;
            Elitra Pharmaceuticals, Inc. (US)
FEATURES              Location/Qualifiers
     source           1. .1239
                      /organism="Candida albicans"
                      /mol_type="unassigned DNA"
                      /db_xref="taxon:5476"
ORIGIN

  Query Match            99.7%;  Score 1235.8;  DB 6;  Length 1239;
  Best Local Similarity  99.8%;  Pred. No. 5.4e-190;
  Matches 1237;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTATTGGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 180

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
```

```
Qy    301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
Db    301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Qy    361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
Db    361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420

Qy    421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
Db    421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480

Qy    481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
Db    481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540

Qy    541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
Db    541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600

Qy    601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
Db    601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660

Qy    661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
Db    661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720

Qy    721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
Db    721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780

Qy    781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
Db    781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840

Qy    841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
Db    841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900

Qy    901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
Db    901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960

Qy    961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
Db    961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020

Qy   1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
Db   1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080

Qy   1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
Db   1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140

Qy   1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
Db   1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200

Qy   1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239
Db   1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239
```

```
RESULT 4
YSCNOFEAT
LOCUS       YSCNOFEAT            1560 bp    DNA     linear   PLN 27-APR-1993
DEFINITION  Saccharomyces cerevisiae transcription factor IIIA and RNA
            polymerase subunit RPB6 genes.
ACCESSION   M90638
VERSION     M90638.1  GI:172046
KEYWORDS    RNA polymerase; transcription factor IIIA.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1560)
  AUTHORS   Woychik,N.A. and Young,R.A.
  TITLE     Genes encoding transcription factor IIIA and the RNA polymerase
            common subunit RPB6 are divergently transcribed in Saccharomyces
            cerevisiae
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89 (9), 3999-4003 (1992)
  MEDLINE   92237295
   PUBMED   1570325
COMMENT     Original source text: Saccharomyces cerevisiae (strain S288C) DNA.
FEATURES             Location/Qualifiers
     source          1..1560
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="S288C"
                     /db_xref="taxon:4932"
ORIGIN      chromosome 14.

  Query Match             17.5%;  Score 217.4;  DB 8;  Length 1560;
  Best Local Similarity   59.4%;  Pred. No. 1e-25;
  Matches  409;  Conservative    0;  Mismatches  271;  Indels    9;  Gaps    2;

Qy         62 GTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCAT 121
Db        368 GGCCAAAGACATATTTCTGTGACTATGATGGCTGTGATAAGGCATTTACAAGACCTTCAA 427

Qy        122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACG 181
Db        428 TTTTGACTGAACACCAATTAAGCGTACATCAGGGTTTA---AGAGCATTTCAGTGTGATA 484

Qy        182 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCG 241
Db        485 AGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTATACGCATTCTG 544

Qy        242 AAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGA 301
Db        545 ATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCGCCAGCAACTGA 604

Qy        302 AAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAG 361
Db        605 AGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGGATGCAACCTCC 664

Qy        362 CATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAA 421
Db        665 GATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTTACATAAACTAA 724

Qy        422 CGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAA 481
Db        725 CCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAATCACATTTCTA 784

Qy        482 AACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAA 535
Db        785 AACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGGTTGTTGCAAAG 844

Qy        536 ATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAAT 595
Db        845 AGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCCTAAATTAAAAT 904

Qy        596 GTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTC 655
Db        905 GTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCACATGATTATTC 964

Qy        656 ATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTG 715
Db        965 ATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGATATGTCTTTTT 1024

Qy        716 CAAAGAAAAATGAATTAGTTGAACATTAT 744
Db       1025 CTAGAAAACATGATCTTCTCACGCATTAT 1053
```

```
RESULT 5
YSCTFIIIA
LOCUS       YSCTFIIIA               1739 bp    DNA     linear   PLN 13-SEP-1996
DEFINITION  Saccharomyces cerevisiae transcription factor IIIA (TFIIIA) gene,
            complete cds.
ACCESSION   M80611
VERSION     M80611.1  GI:172902
KEYWORDS    transcription factor IIIA; zinc-finger protein; zinc-finger
            transcription factor.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 1739)
  AUTHORS   Archambault,J., Milne,C.A., Schappert,K.T., Baum,B., Friesen,J.D.
            and Segall,J.
  TITLE     The deduced sequence of the transcription factor TFIIIA from
            Saccharomyces cerevisiae reveals extensive divergence from Xenopus
            TFIIIA
  JOURNAL   J. Biol. Chem. 267 (5), 3282-3288 (1992)
  MEDLINE   92147684
   PUBMED   1737784
FEATURES             Location/Qualifiers
     source          1. .1739
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /strain="W303-1A"
                     /db_xref="taxon:4932"
                     /chromosome="XVI"
     gene            234. .1523
                     /gene="TFIIIA"
     CDS             234. .1523
                     /gene="TFIIIA"
                     /note="zinc finger containing protein"
                     /codon_start=1
                     /product="transcription factor IIIA"
                     /protein_id="AAB08014.1"
                     /db_xref="GI:172903"
                     /translation="MGGEVLNNEGMPLAELKQETIPISRSESSESLNSLTSTRSSSSN
                     RPKTYFCDYDGCDKAFTRPSILTEHQLSVHQGLRAFQCDKCAKSFVKKSHLERHLYTH
                     SDTKPFQCSYCGKGVTTRQQLKRHEVTHTKSFICPEEGCNLRFYKHPQLRAHILSVHL
                     HKLTCPHCNKSFQRPYRLRNHISKHHDPEVENPYQCTFAGCCKEFRIWSQLQSHIKND
                     HPKLKCPICSKPCVGENGLQMHMIIHDDSLVTKNWKCHICPDMSFSRKHDLLTHYGSI
                     HTEEDIPLELKYKISDIQQLVQDHGVQLGNSKHSNEQDEEKISNRLRKRRKLTENNNV
                     EFLQNEVDLEKRLESGENGLNLLLNTVGRKYRCFYNNCSRTPKTKEKYEKHIDKHKVH
                     ELKLKILQEKEENKTLVDQNHKEPFIIQKETQSAGDK"
ORIGIN

  Query Match           17.5%;  Score 217.4;  DB 8;  Length 1739;
  Best Local Similarity  59.4%;  Pred. No. 9.9e-26;
  Matches  409;  Conservative    0;  Mismatches  271;  Indels    9;  Gaps    2;

Qy         62 GTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCAT 121
Db        367 GGCCAAAGACATATTTCTGTGACTATGATGGCTGTGATAAGGCATTTACAAGACCTTCAA 426

Qy        122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACG 181
Db        427 TTTTGACTGAACACCAATTAAGCGTACATCAGGGTTTA---AGAGCATTTCAGTGTGATA 483

Qy        182 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCG 241
Db        484 AGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTATACGCATTCTG 543

Qy        242 AAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGA 301
Db        544 ATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCGCCAGCAACTGA 603

Qy        302 AAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAG 361
Db        604 AGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGGATGCAACCTCC 663

Qy        362 CATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAA 421
Db        664 GATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTTACATAAACTAA 723

Qy        422 CGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAA 481
Db        724 CCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAATCACATTTCTA 783

Qy        482 AACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAA 535
Db        784 AACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGGTTGTTGCAAAG 843

Qy        536 ATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAAT 595
Db        844 AGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCCTAAATTAAAAT 903

Qy        596 GTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTC 655
Db        904 GTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCACATGATTATTC 963

Qy        656 ATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTG 715
Db        964 ATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGATATGTCTTTTT 1023

Qy        716 CAAAGAAAAATGAATTAGTTGAACATTAT 744
Db       1024 CTAGAAAACATGATCTTCTCACGCATTAT 1052


RESULT 6
YSCP9677/c
LOCUS       YSCP9677                37497 bp    DNA     linear   PLN 01-AUG-1997
DEFINITION  Saccharomyces cerevisiae chromosome XVI cosmid 9677.
ACCESSION   U25841 U00094
VERSION     U25841.1  GI:786295
KEYWORDS    .
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 37497)
  AUTHORS   Johnston,M., Andrews,S., Brinkman,R., Cooper,J., Ding,H., Du,Z.,
            Favello,A., Fulton,L., Gattung,S., Greco,T., Kirsten,J., Kucaba,T.,
            Hallsworth,K., Hawkins,J., Hillier,L., Jier,M., Johnson,D.,
            Johnston,L., Langston,Y., Latreille,P., Le,T., Mardis,E.,
            Menezes,S., Miller,N., Nhan,M., Pauley,A., Peluso,D., Rifken,L.,
            Riles,L., Taich,A., Trevaskis,E., Vignati,D., Wilcox,L.,
            Wohldman,P., Vaudin,M., Wilson,R. and Waterston,R.
  TITLE     The sequence of a portion of the right arm of Saccharomyces
            cerevisiae chromosome XVI
  JOURNAL   Unpublished (1995)
REFERENCE   2  (bases 1 to 37497)
  AUTHORS   Miller,N.
  TITLE     The sequence of S. cerevisiae cosmid 9677
  JOURNAL   Unpublished (1995)
REFERENCE   3  (bases 1 to 37497)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-APR-1995) Robert Waterston
COMMENT     Submitted by:
                      Genome Sequencing Center
                      Department of Genetics, Washington University,
                      St. Louis, MO  63110, USA
                      e-mail: mj@sequencer.wustl.edu

            NEIGHBORING COSMID INFORMATION:
            This sequence includes nucleotides 1-37497 of cosmid 9677.  The
            cosmid on the left is 9705.
FEATURES             Location/Qualifiers
     source          1. .37497
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
```

```
                /strain="S288C (AB972)"
                /db_xref="taxon:4932"
                /map="XVI"
                /note="baker's yeast"
gene            complement(2366. .3655)
                /gene="TFC2"
CDS             complement(2366. .3655)
                /gene="TFC2"
                /note="P9677.9"
                /codon_start=1
                /product="Transcription factor TFIIIA (PIR accession
                number S20050)"
                /protein_id="AAB64615.1"
                /db_xref="GI:786305"
                /translation="MGGEVLNNEGMPLAELKQETIPISRSESSESLNSLTSTRSSSSN
                RPKTYFCDYDGCDKAFTRPSILTEHQLSVHQGLRAFQCDKCAKSFVKKSHLERHLYTH
                SDTKPFQCSYCGKGVTTRQQLKRHEVTHTKSFICPEEGCNLRFYKHPQLRAHILSVHL
                HKLTCPHCNKSFQRPYRLRNHISKHHDPEVENPYQCTFAGCCKEFRIWSQLQSHIKND
                HPKLKCPICSKPCVGENGLQMHMIIHDDSLVTKNWKCHICPDMSFSRKHDLLTHYGSI
                HTEEDIPLELKYKISDIQQLVQDHGVQLGNSKHSNEQDEEKISNRLRKRRKLTENNNV
                EFLQNEVDLEKRLESGENGLNLLLNTVGRKYRCFYNNCSRTFKTKEKYEKHIDKHKVH
                ELKLKILQEKEENKTLVDQNHKEPFIIQKETQSAGDK"
gene            3890. .4433
                /gene="RPB6"
CDS             join(3890. .3909,3986. .4433)
                /gene="RPB6"
                /note="P9677.8"
                /codon_start=1
                /product="23 kD polypeptide of RNA Polymerases I, II, and
                III (Swiss Prot. accession number P20435)"
                /protein_id="AAB64616.1"
                /db_xref="GI:786304"
                /translation="MSDYEEAFNDGNENFEDFDVEHFSDEETYEEKPQFKDGETTDAN
                GKTIVTGGNGPEDFQQHEQIRRKTLKEKAIPKDQRATTPYMTKYERARILGTRALQIS
                MNAPVFVDLEGETDPLRIAMKELAEKKIPLVIRRYLPDGSFEDWSVEELIVDL"
gene            complement(4621. .5112)
                /gene="P9677.10"
CDS             complement(4621. .5112)
                /gene="P9677.10"
                /note="Similar to calmodulins (e.g., Swiss Prot. accession
                number P13868) and calmodulin-related proteins (e.g.,
                Swiss Prot. accession number P27482) from several
                organism."
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAB64617.1"
                /db_xref="GI:786306"
                /translation="MDHSESLTFNQLTQDYINKLKDAFQMLDEDEDGLISRGDLTKIY
                ATLGKTLTDEEWSKMVPDNDTSTAEVGEEGVSFPIFLSIMGKNLSQFPEREELEESLK
                AIGRGHDLNVPLNEVIDSLKEAGFENPEEEFAKLFKLFTTNQQATEERTFRGKLFLDS
                ITD"
gene            5297. .9595
                /gene="SKI3"
CDS             5297. .9595
                /gene="SKI3"
                /note="P9677.7"
                /codon_start=1
                /product="Contains 8 copies of the TPR domain"
                /protein_id="AAB64618.1"
                /db_xref="GI:786303"
                /translation="MSDIKQLLKEAKQELTNRDYEETIEISEKVLKLDPDNYFAHIFL
                GKALSSLPASNNVSSNRNLERATNHYVSAAKLVPDNLLAWKGLFLLFRTTEVVPDILS
                YDEYFDLCGQYADALLKQEQSQVELINDIKLLKKTHPDCQKAFYQHLKPGSLMAETIG
                RHLSTPQDALLNLIKILSNIETTEIGKTLSQNRLKLKASDPDYQIKLNSFSWEIIKNS
                EIDQLYNQLVNILADDQKRSEIENQWLEYRIKVLKSMPLDVKKDFFTKVKEMVEDMVL
                VNHQSLLAWQKYFEWTDYEDLDNMDAPLIIKYFKKFPKDPLAMILYSWLSSKLSKYDI
                KSLESANKPPEGHKKTEKETDIKDVDETNEDEVKDRVEDEVKDRVEDEVKDQDEEAKE
                DEEEDLDDIEIGLLEEEVVTVLTENIVKCKNNILAHRILCQYYLLTKEYEAALPYIKN
                GISLIAYNIKDLGVHLPLTKREFSLDLATVYTYVDAPKDHNAALKLYDNILSGDFSNI
                QAKMGKGIIFIERKNWKDAMTLLTQVHEQSPNNLEVLSELSWSKAHMGYMDEALAGLD
                TVIKGIKGMDLRSIDFRALNLWRQAKVYIMKHASINDAKQENVKCAFKLLIQSIKILD
                TFAPGFSTLGDIYCHYYKDHLRAFKCYFKAFDLDAGDYTAAKYITETYASKPNWQAAS
                SIASRLIKGEKAKAELRSNNWPFRVVGIAHLEKQEESDSIEWFQSALRVDPNDVESWV
```

```
                GLGQAYHACGRIEASIKVFDKAIQLRPSHTFAQYFKAISLCDVGEYLESLDILEKVCQ
                EAATEESFQIGLVEVLMRCSLDLYSQGFLLKSVSIAKDTIERIKIIISELKCENQQVW
                IYLSQVLRLFIWIESKVDTLPVESLVSIFENSQFSGSEEIDSVDNIKIDTLLDSTTDD
                NVSIACKFLILASKYSVSDQKFTDIAGTVRASYWYNIGISELTAFITLKEPQYRDAAI
                FAFKKSIQLQSNTSETWIGLGIATMDINFRVSQHCFIKATALEPKATNTWFNLAMLGL
                KKKDTEFAQQVLNKLQSLAPQDSSPWLGMALILEEQGDIIGSSKLFAHSFILSNGRSK
                AAQFMYAKNVLENHINNGDDERDIETVEKLTTASIALEQFFKKSPDSQFALQCALLTL
                ERLHHYENANELANRLIGILEKKFEKTQDERELFNFAIIKGQFARIHLGLGNFELSIE
                NADLSQGIISESSDEKSMKTKISNHICLGLSYFFLNDFDQTLNQFQELLSISKDSKHL
                VVLIAKVLYDVGESDTKEIALQELTBYIATSGADLLVTLTIAAMSILDDKREDLSIIL
                EELKALPLSKQIIDKHKDAPYLIEEITKRLYRNDTGKQVWQRSAYFFPNNLKVWERLD
                KNIQRRIASNGQNKVTAEEMSKLYCESKNLRSIQRGMFLCPWNVTAVKALNECF"
gene            complement(9710. .11674)
                /gene="RPC3"
CDS             complement(9710. .11674)
                /gene="RPC3"
                /note="P9677.11"
                /codon_start=1
                /product="74 kD subunit of RNA Polymerase III (Swiss Prot.
                accession number P32349)"
                /protein_id="AAB64619.1"
                /db_xref="GI:786307"
                /translation="MDELLGEALSAENQTGESTVESEKLVTPEDVMTISSLEQRTLNP
                DLFLYKELVKAHLGERAASVIGMLVALGRLSVRELVEKIDGMDVDSVKTTLVSLTQLR
                CVKYLQETAISGKKTTYYYYNEEGIHILLYSGLIIDEIITQMRVNDEEEHKQLVAEIV
                QNVISLGSLTVEDYLSSVTSDSMKYTISSLFVQLCEMGYLIQISKLHYTPIEDLWQFL
                YEKHYKNIPRNSPLSDLKKRSQAKMNAKTDFAKIINKPNELSQILTVDPKTSLRIVKP
                TVSLTINLDRFMKGRRSKQLINLAKTRVGSVTAQVYKIALRLTEQKSPKIRDPLTQTG
                LLQDLEEAKSFQDEAELVEEKTPGLTFNAIDLARHLPAELDLRGSLLSRKPSDNKKRS
                GSNAAASLPSKKLKTEDGFVIPALPAAVSKSLQESGDTQEEDEEEEDLDADTEDPHSA
                SLINSHLKILASSNFPFLNETKPGVYYVPYSKLMPVLKSSVYEYVIASTLGPSAMRLS
                RCIRDNKLVSEKIINSTALMKEKDIRSTLASLIRYNSVEIQEVPRTADRSASRAVFLF
                RCKETHSYNFMRQNLEWNMANLLFKKEKLKQENSTLLKKANRDDVKGRENELLLPSEL
                NQLKMVNERELNVFARLSRLLSLWEVFQMA"
gene            12014. .13120
                /gene="QCR2"
CDS             12014. .13120
                /gene="QCR2"
                /note="P9677.6"
                /codon_start=1
                /product="Ubiquinol-cytochrome C reductase core protein 2"
                /protein_id="AAB64620.1"
                /db_xref="GI:786302"
                /translation="MLSAARLQFAQGSVRRLTVSARDAPTKISTLAVKVHGGSRYATK
                DGVAHLLNRFNFQNTNTRSALKLVRESELLGGTFKSTLDREYITLKATFLKDDLPYYV
                NALADVLYKTAFKPHELTESVLPAARYDYAVAEQCPVKSAEDQLYAITFRKGLGNPLL
                YDGVERVSLQDIKDFADKVYTKENLEVSGENVVEADLKRFVDESLLSTLPAGKSLVSK
                SEPKFFLGEENRVRFIGDSVAAIGIPVNKASLAQYEVLANYLTSALSELSGLISSAKL
                DKFTDGGLFTLFVRDQDSAVVSSNIKKIVADLKKGKDLSPAINYTKLKNAVQNESVSS
                PIELNFDAVKDFKLGKFNYVAVGDVSNLPYLDEL"
gene            14493. .15410
                /gene="P9677.5"
CDS             14493. .15410
                /gene="P9677.5"
                /note="Belongs to the MIP family of transmembrane channel
                proteins."
                /codon_start=1
                /evidence=not_experimental
                /protein_id="AAB64621.1"
                /db_xref="GI:786301"
                /translation="MSSNDSNDTDKQHTRLDPTGVDDAYIPPEQPETKHHRFKISRDT
                LRDHFIAAVGEFCGTFMFLWCAYVICNVANHDVALVAAPDGSHPGQLIMIAIGFGFSV
                MFSIWCFAGVSGGALNPAMSLSLCLARAVSPTRCVVMWVSQIVAGMAAGGAASAMTPG
                EVLFANSLGLGCSRTRGLFLEMFGTAILCLTVLMTAVEKRETNFMAALPIGISLFIAH
                VALTAYTGTGVNPARSLGAAVAARYFPHYHWIYWIGTLLGSILAWSVWQLLQILDYTT
                YVTAEKAASTKEKAQKKGETSSSSAVAEV"
gene            complement(15542. .16012)
                /gene="P9677.12"
CDS             complement(15542. .16012)
                /gene="P9677.12"
                /note="Similar in C-terminus to Gentamicin
                3'-acetyltransferase of P. aeruginosa (Swiss Prot.
                accession number P23181)"
```

```
                     /codon_start=1
                     /evidence=not_experimental
                     /protein_id="AAB64622.1"
                     /db_xref="GI:786308"
                     /translation="MSNTSEDNITVRFVTENDKEGWQRLWKSYQDFYEVSFPDDLDDF
                     NFGRFLDPNIKMWAAVAVESSSEKIIGMINFFNHMTTWDFKDKIYINDLYVDENSRVK
                     GAGGKLIQFVYDEADKLGTPSVYWCTDESNHRAQLLYVKVGYKAPKILYKRKGY"
     gene            complement(16937. .19570)
                     /gene="P9677.13"
     CDS             complement(16937. .19570)
                     /gene="P9677.13"
                     /note="Similar to S. cerevisiae hypothetical protein
                     HRD799 (PIR accession number S45161) and S. pombe

  Query Match             17.5%;  Score 217.4;  DB 8;  Length 37497;
  Best Local Similarity   59.4%;  Pred. No. 4.8e-26;
  Matches  409;  Conservative    0;  Mismatches  271;  Indels    9;  Gaps    2;

Qy         62 GTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCAT 121
Db       3522 GGCCAAAGACATATTTCTGTGACTATGATGGCTGTGATAAGGCATTTACAAGACCTTCAA 3463

Qy        122 TATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACG 181
Db       3462 TTTTGACTGAACACCAATTAAGCGTACATCAGGGTTTA---AGAGCATTTCAGTGTGATA 3406

Qy        182 ATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCG 241
Db       3405 AGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTATACGCATTCTG 3346

Qy        242 AAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGA 301
Db       3345 ATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCGCCAGCAACTGA 3286

Qy        302 AAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAG 361
Db       3285 AGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGGATGCAACCTCC 3226

Qy        362 CATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAA 421
Db       3225 GATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTTACATAAACTAA 3166

Qy        422 CGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAA 481
Db       3165 CCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAATCACATTTCTA 3106

Qy        482 AACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAA 535
Db       3105 AACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGGTTGTTGCAAAG 3046

Qy        536 ATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAAT 595
Db       3045 AGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCCTAAATTAAAAT 2986

Qy        596 GTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTC 655
Db       2985 GTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCACATGATTATTC 2926

Qy        656 ATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTG 715
Db       2925 ATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGATATGTCTTTTT 2866

Qy        716 CAAAGAAAAATGAATTAGTTGAACATTAT 744
Db       2865 CTAGAAAACATGATCTTCTCACGCATTAT 2837


RESULT 7
YSCRP026A/c
LOCUS       YSCRP026A               2133 bp    DNA     linear   PLN 27-APR-1993
DEFINITION  S.cerevisiae RNA polymerase II sixth subunit (RP026) gene, complete
            cds.
ACCESSION   M33924
VERSION     M33924.1  GI:172452
KEYWORDS    RNA polymerase II; c-myc proto-oncogene; transmembrane protein;
            tyrosine kinase.
SOURCE      Saccharomyces cerevisiae (baker's yeast)
  ORGANISM  Saccharomyces cerevisiae
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 2133)
  AUTHORS   Archambault,J., Schappert,K.T. and Friesen,J.D.
  TITLE     A suppressor of an RNA polymerase II mutation of Saccharomyces
            cerevisiae encodes a subunit common to RNA polymerases I, II, and
            III
  JOURNAL   Mol. Cell. Biol. 10 (12), 6123-6131 (1990)
  MEDLINE   91061718
   PUBMED   2247052
COMMENT     Original source text: S.cerevisiae single-copy plasmid DNA, clone
            2.
            Draft entry and computer-readable sequence for [Unpublished (1990)]
            kindly submitted
            by J.D.Friesen, 01-MAY-1990.
               The Hospital for Sick Children
               555 University Avenue
               Toronto, Ontario, CANADA M5G 1X8
               e-mail: jim@sickkids.toronto.edu.
FEATURES             Location/Qualifiers
     source          1. .2133
                     /organism="Saccharomyces cerevisiae"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4932"
     CDS             join(1026. .1045,1122. .1569)
                     /note="RNA polymerase II sixth subunit (RP026)"
                     /codon_start=1
                     /protein_id="AAA34989.1"
                     /db_xref="GI:172453"
                     /translation="MSDYEEAFNDGNENFEDFDVEHFSDEETYEEKPQFKDGETTDAN
                     GKTIVTGGNGPEDFQQHEQIRRKTLKEKAIPKDQRATTPYMTKYERARILGTRALQIS
                     MNAPVFVDLEGETDPLRIAMKELAEKKIPLVIRRYLPDGSFEDWSVEELIVDL"
     exon            <1026. .1045
                     /note="RNA polymerase II sixth subunit (RP026)"
                     /number=1
     intron          1046. .1121
                     /note="RP026 intron A"
     exon            1122. .>1569
                     /note="RNA polymerase II sixth subunit (RP026)"
                     /number=2
ORIGIN      Chromosome XVI.

  Query Match             16.1%;  Score 199.2;  DB 8;  Length 2133;
  Best Local Similarity   61.4%;  Pred. No. 8.2e-23;
  Matches  340;  Conservative    0;  Mismatches  208;  Indels    6;  Gaps    1;

Qy        171 TACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGT 230
Db        556 TTCAGTGTGATAGTGTGCAAAATCCTTCGTTAAAAAGAGTCACTTAGAGAGACACTTGTA 497

Qy        231 ATCACATTCCGAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCG 290
Db        496 TACGCATTCTGATACGAAACCATTCCAATGTTCTTATTGTGGAAAAGGAGTGACGACTCG 437

Qy        291 ACAACACTTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAA 350
Db        436 CCAGCAACTGAAGCGACACGAAGTAACGCATACCAAATCTTTCATTTGTCCAGAGGAAGG 377

Qy        351 TTGTCAAGAAGCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGA 410
Db        376 ATGCAACCTCCGATTCTACAAGCATCCACAATTAAGGGCACATATTTTATCTGTTCATTT 317

Qy        411 AAAAACATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACA 470
Db        316 ACATAAACTAACCTGTCCACACTGCAATAAAAGCTTTCAGAGGCCATATAGGCTAAGAAA 257

Qy        471 ACATAAATTAAAACATCA------TGGTGGATCTCCTGCTTATCAATGTGATCATCCTGG 524
```

```
Db        256 TCACATTTCTAAACATCACGATCCTGAGGTAGAAAATCCTTACCAATGTACTTTTGCTGG 197

Qy        525 TTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCC 584

Db        196 TTGTTGCAAAGAGTTTCGGATATGGTCACAATTGCAATCGCATATAAAAAATGATCATCC 137

Qy        585 AAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACA 644

Db        136 TAAATTAAAATGTCCTATTTGTAGCAAACCATGTGTGGGGGAAAATGGTTTACAAATGCA 77

Qy        645 TATGTTAAGTCATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGT 704

Db         76 CATGATTATTCATGACGACTCACTAGTAACCAAAAATTGGAAGTGTCATATATGTCCTGA 17

Qy        705 GGGGAAATTTGCAA 718

Db         16 TATGTCTTTTTCTA 3
```

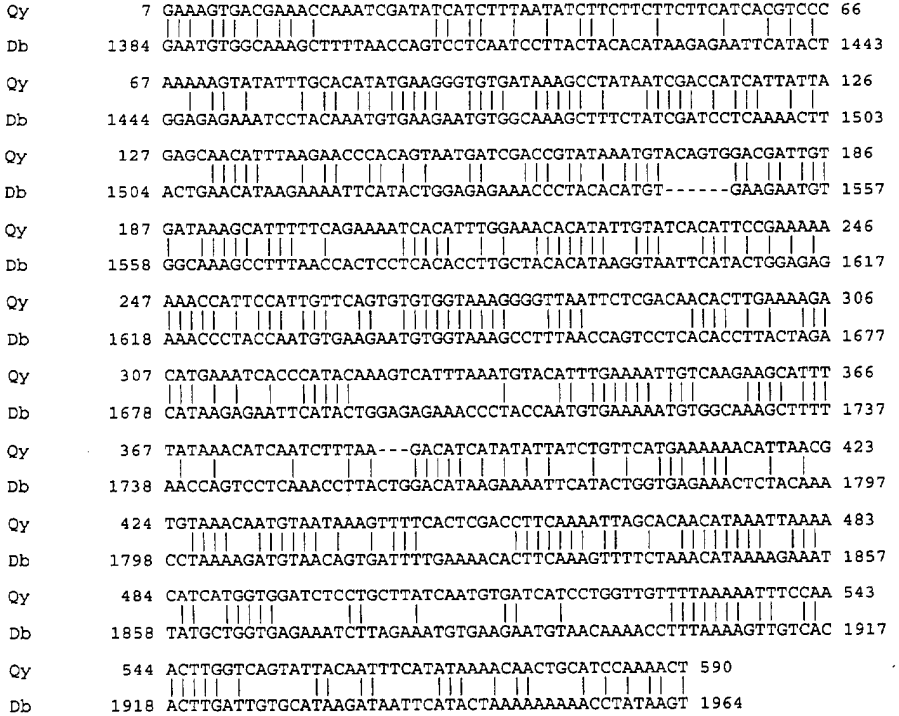
RESULT 8
CNS06ING
LOCUS CNS06ING 797 bp DNA linear STS 30-NOV-2001
DEFINITION T7 end of clone AS0AA022B07 of library AS0AA from strain CLIB 533 of Saccharomyces bayanus, sequence tagged site.
ACCESSION AL400514
VERSION AL400514.1 GI:12156635
KEYWORDS STS.
SOURCE Saccharomyces bayanus
ORGANISM Saccharomyces bayanus
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 797)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 797)
AUTHORS Bon,E., Neuveglise,C., Casaregola,S., Artiguenave,F., Wincker,P., Aigle,M. and Durrens,P.
TITLE Genomic exploration of the hemiascomycetous yeasts: 5. Saccharomyces bayanus var. uvarum
JOURNAL FEBS Lett. 487 (1), 37-41 (2000)
MEDLINE 20584715
PUBMED 11152880
REFERENCE 3 (bases 1 to 797)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This STS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES Location/Qualifiers
source 1. .797
/organism="Saccharomyces bayanus"
/mol_type="genomic DNA"
/strain="CLIB 533"
/variety="uvarum"
/db_xref="taxon:4931"
/clone="AS0AA022B07"
/clone_lib="AS0AA"
/note="end : T7"
misc_feature <2. .>796
/note="similar to Saccharomyces cerevisiae ORF YPR186c [PZF1 ; TFIIIA (transcription initiation factor)]"
/evidence=not_experimental
ORIGIN

```
  Query Match             15.6%;  Score 193;  DB 11;  Length 797;
  Best Local Similarity   59.3%;  Pred. No. 1e-21;
  Matches  347;  Conservative    1;  Mismatches  231;  Indels    6;  Gaps    1;

Qy        178 GACGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACAT 237

Db         83 GACAATTGTGCAAAATCCTTTGTTAARAAGAGTCATCTAGAGAGGCACTTGTTTTCGCAT 142

Qy        238 TCCGAAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACAC 297

Db        143 TCTGATACGAAGCCATTCCAATGTTCTTATTGTGGCAAGGGGGTGACAACTCGACAGCAA 202

Qy        298 TTGAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAA 357

Db        203 TTGAAGCGCCACGAAGTTACACATACCAAATCTTTTGTTTGTCCGGAAGATGGATGTGAT 262

Qy        358 GAAGCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACA 417

Db        263 CTCCGATTTTATAAGCACCCTCAATTAAGGGCTCATATTCTATCTGTCCACCTAGACAAA 322

Qy        418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477

Db        323 TTAACCTGTCCACAATGTAGTAAAARTTTTCAGAGACCCTACAGGCTAAGGAATCATATC 382

Qy        478 TTAAAACATCATGGT------GGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTT 531

Db        383 TCTAAACATCATGATCCTGAAGTGGTAAATCCGTATCAATGTACTTTCGCTGGCTGTTGT 442

Qy        532 AAAAATTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTT 591

Db        443 ACAGAGTACCGTATATGGTCACAGTTGCAATCACATATCAAAAATGACCATCCTAAATTG 502

Qy        592 AAATGTCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTA 651

Db        503 AAATGCCCCATTTGCAACAAGCCTTGTGTAGGGGAAAATGGTCTACAAATGCATATGATC 562

Qy        652 AGTCATGATGATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAA 711

Db        563 ATTCATGATGAATCTTTGGTAACTAAGAACTGGAAATGTCATATTTGTTCTAATACATCT 622

Qy        712 TTTGCAAAGAAAAATGAATTAGTTGAACATTATAATATCTTCCAT 756

Db        623 TTCTTGAGAAAGCACGACCTCCTTACTCACTACCAGATGACGCAT 667
```

RESULT 9
AC138126/c
LOCUS AC138126 169063 bp DNA linear PRI 17-DEC-2002
DEFINITION Homo sapiens chromosome 19 clone RP11-274A19, complete sequence.
ACCESSION AC138126
VERSION AC138126.1 GI:27151357
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169063)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 169063)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission

```
JOURNAL     Submitted (17-DEC-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
FEATURES             Location/Qualifiers
     source          1..169063
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="RP11-274A19"
ORIGIN

  Query Match             8.2%;  Score 101;  DB 9;  Length 169063;
  Best Local Similarity   50.8%;  Pred. No. 2e-07;
  Matches  298;  Conservative    0;  Mismatches  280;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
Db      80134 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 80075

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
Db      80074 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 80015

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
Db      80014 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 79961

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
Db      79960 GGCAAAGCCTTTAACCATTCCTCACATCTTGCTACACATAAGGTAATTCATACTGGAGAG 79901

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
Db      79900 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 79841

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
Db      79840 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 79781

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
Db      79780 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 79721

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
Db      79720 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 79661

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
Db      79660 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 79601

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
Db      79600 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTACAAGT 79554

RESULT 10
AC073544/c
LOCUS       AC073544               177299 bp    DNA     linear   PRI 19-JUN-2002
DEFINITION  Homo sapiens chromosome 19 clone RP11-359H18, complete sequence.
ACCESSION   AC073544
VERSION     AC073544.4  GI:21465367
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 177299)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 177299)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 177299)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-JUN-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jun 19, 2002 this sequence version replaced gi:13699752.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.2.
            NOTE:  BACTERIAL TRANSPOSONS excised at 14620 and 98099.
FEATURES             Location/Qualifiers
     source          1..177299
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="RP11-359H18"
ORIGIN

  Query Match             8.2%;  Score 101;  DB 9;  Length 177299;
  Best Local Similarity   50.8%;  Pred. No. 2e-07;
  Matches  298;  Conservative    0;  Mismatches  280;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
Db     163740 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 163681

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
Db     163680 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 163621

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
Db     163620 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 163567

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
Db     163566 GGCAAAGCCTTTAACCATTCCTCACATCTTGCTACACATAAGGTAATTCATACTGGAGAG 163507

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
Db     163506 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 163447

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
Db     163446 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 163387

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
Db     163386 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 163327

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
Db     163326 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 163267

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
Db     163266 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 163207

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
```

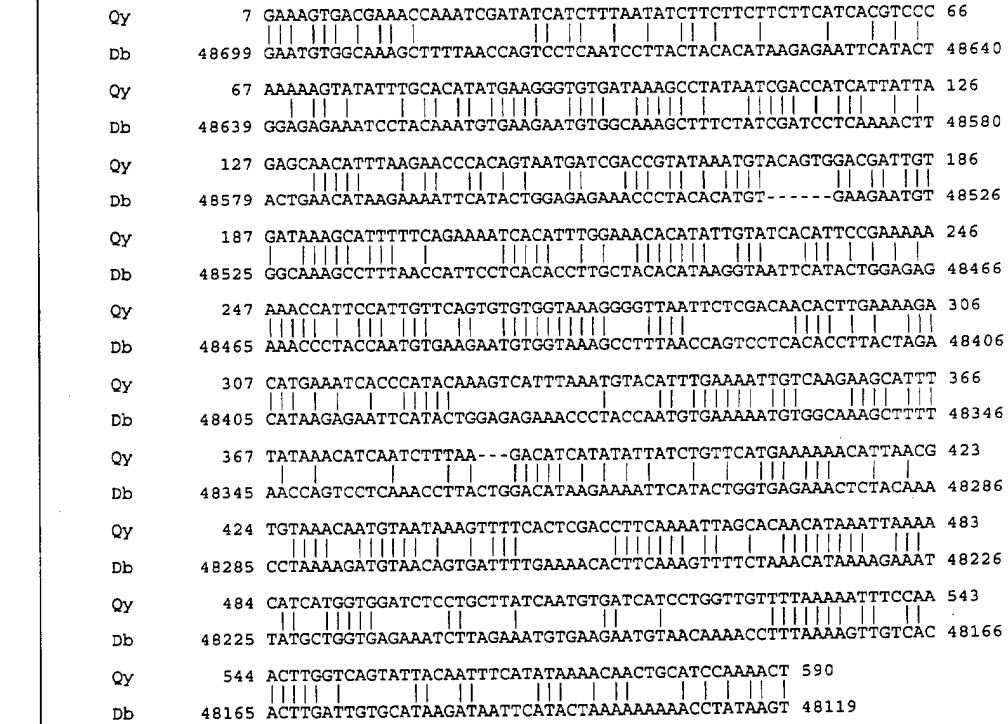
```
Db     163206 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTACAAGT 163160


RESULT 11
AX714042
LOCUS       AX714042               2132 bp    DNA     linear   PAT 15-APR-2003
DEFINITION  Sequence 726 from Patent EP1293569.
ACCESSION   AX714042
VERSION     AX714042.1  GI:29888970
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Isogai,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
            Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
            Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
            Masuho,Y.
  TITLE     Full-length cDNAs
  JOURNAL   Patent: EP 1293569-A 726 19-MAR-2003;
            Helix Research Institute (JP) ; Research Association for
            Biotechnology (JP)
FEATURES             Location/Qualifiers
     source          1. .2132
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

  Query Match             8.0%;  Score 99.4;  DB 6;  Length 2132;
  Best Local Similarity   50.6%;  Pred. No. 1e-06;
  Matches  297;  Conservative    0;  Mismatches  281;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | ||| |           || ||  |  |  || |    |      | | | |
Db       1384 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 1443

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
               | ||  |    | || || |||||  ||||  |||||  |    ||||| | |||  | |
Db       1444 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 1503

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |||||   | ||  || | |  ||   ||| || | ||||     || || |||
Db       1504 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 1557

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              | ||||| ||| |      |||||  | |  |||||||  |||   ||| | | | | |
Db       1558 GGCAAAGCCTTTAACCACTCCTCACACCTTGCTACACATAAGGTAATTCATACTGGAGAG 1617

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | ||| |||  ||  |||||||||||   ||||          |||| | |  |||
Db       1618 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 1677

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |  |  |||||           |     || |||||| |||    |||| |||
Db       1678 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 1737

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              | |  |     | |  |||||  | | | |  |  |||  |||    |  |
Db       1738 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 1797

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||||  |||||| |  | |||         |||||| ||  |  ||||||||  |||
Db       1798 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 1857

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
              ||  |||||     ||     |     ||  |          ||||||| ||  ||
Db       1858 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 1917

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
              ||||| |    ||  ||     ||| | ||    | | | || |
Db       1918 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTATAAGT 1964


RESULT 12
AK056088
LOCUS       AK056088               2132 bp    mRNA    linear   PRI 01-AUG-2002
DEFINITION  Homo sapiens cDNA FLJ31526 fis, clone NT2RI2000341, moderately
            similar to ZINC FINGER PROTEIN 43.
ACCESSION   AK056088
VERSION     AK056088.1  GI:16551397
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ninomiya,K., Wagatsuma,M., Kanda,K., Kondo,H., Yokoi,T.,
            Kodaira,H., Furuya,T., Takahashi,M., Kikkawa,E., Omura,Y., Abe,K.,
            Kamihara,K., Katsuta,N., Sato,K., Tanikawa,M., Yamazaki,M.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
            Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
            Kikuchi,H., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
            Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
            Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2132)
  AUTHORS   Isogai,T., Otsuki,T. and Sugiyama,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: RAB and
            HRI.
FEATURES             Location/Qualifiers
     source          1. .2132
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RI2000341"
                     /cell_line="NT2"
                     /cell_type="teratocarcinoma"
                     /clone_lib="NT2RI2"
                     /note="cloning vector: pME18SFL3-mRNA from NT2 neuronal
                     precursor cells treated 2-weeks mitotic inhibitor after
                     5-weeks retinoic acid (RA) induction.-majorly NT2 neuron"
     CDS             148. .1878
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="BAB71090.1"
                     /db_xref="GI:16551398"
                     /translation="MVAEPPVICSHFAQHFSPEQNIKDSFQKVTPRRYGKCEHENLQL
                     SKSVDECKVQKGGYNGLNQCLPTTQSKIFQCDKYMKIFHKFSNLNGHKVRHTRKKPFK
                     YKEFGKSFCIFSNLTQHKIICTRVNFYKCEDCGKAFNGSSIFTKHKRIHIGEKSYICE
                     ECGKACNQFTNLTTHKIIYTRDKLYKREECSKAFNLSSHITTHTIIHTGENPYKREEC
                     DKAFNQSSTLTTHKIIHTREKLNEYKECGKAFNQSSHLTRHKIIHTGEKPYKCEECGK
                     AFNQSSHLTRHKIIHTGEKPYKCEECGKAFRQSSHLTTHKIIHTGEKPYKCEECGKAF
                     NKSSHLTRHKSIHTGEKPYQCEKCGKASNQSSNLTEHKNIHTEEKPYKCEECGKAFNQ
                     FSNLTTHKRIHTGEKPYKCEECGKAFNQSSILTTHKRIHTGEKSYKCEECGKAFYRSS
                     KLTEHKKIHTGEKPYTCEECGKAFNHSSHLATHKVIHTGEKPYQCEECGKAFNQSSHL
                     TRHKRIHTGEKPYQCEKCGKAFNQSSNLTGHKKIHTGEKLYKPKRCNSDFENTSKFSK
                     HKRNYAGEKS"
ORIGIN

  Query Match             8.0%;  Score 99.4;  DB 9;  Length 2132;
  Best Local Similarity   50.6%;  Pred. No. 1e-06;
  Matches  297;  Conservative    0;  Mismatches  281;  Indels    9;  Gaps    2;
```



RESULT 13
AC099500/c
LOCUS AC099500 138627 bp DNA linear PRI 29-MAY-2002
DEFINITION Homo sapiens chromosome 19 clone RP11-209J6, complete sequence.
ACCESSION AC099500
VERSION AC099500.2 GI:21240690
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 138627)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 138627)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 138627)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (29-MAY-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On May 29, 2002 this sequence version replaced gi:16930916.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www-shgc.stanford.edu
Quality: Phrap Quality >=40 100% of Sequence;

Estimated Total Number of Errors is 0.1.
NOTE: This insert is not the entire sequence of the clone (entire sequence is 153.8kb). It is clipped at the overlap with AC008981.
The number of bases overlapped is 27192.
FEATURES Location/Qualifiers
source 1. .138627
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-209J6"
ORIGIN

Query Match 8.0%; Score 99.4; DB 9; Length 138627;
Best Local Similarity 50.6%; Pred. No. 3.9e-07;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;



RESULT 14
AC138469/c
LOCUS AC138469 169500 bp DNA linear HTG 08-JAN-2003
DEFINITION Homo sapiens chromosome 19 clone RP11-189C24, *** SEQUENCING IN PROGRESS ***, 5 ordered pieces.
ACCESSION AC138469
VERSION AC138469.1 GI:27544967
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 169500)
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission

```
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169500)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2003) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        1    75796: contig of 75796 bp in length
            *    75797    75896: gap of unknown length
            *    75897   138071: contig of 62175 bp in length
            *   138072   138171: gap of unknown length
            *   138172   157899: contig of 19728 bp in length
            *   157900   157999: gap of unknown length
            *   158000   166058: contig of 8059 bp in length
            *   166059   166158: gap of unknown length
            *   166159   169500: contig of 3342 bp in length.
            -------Genome Center
                        Center: Joint Genome Institute
                        Center Code: JGI
                        Web site: http://www.jgi.doe.gov
                        -------
                        Project Information
                        Center Project Name: 479013
                        Center clone name: RP11-189C24
                        -------
                        Summary Statistics
                        Consensus quality: 165586 bases at least Q40
                        Consensus quality: 165891 bases at least Q30
                        Consensus quality: 166156 bases at least Q20
                        Estimated insert size: 169095000; agarose-fp estimation
                        Estimated insert size: 166687; sum-of-contigs
            estimation
                        Quality coverage:  0 in Q20 bases; agarose-fp
            estimation
                        Quality coverage:  1 in Q20 bases; sum-of-contigs
            estimation.
FEATURES              Location/Qualifiers
     source           1. .169500
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /chromosome="19"
                      /clone="RP11-189C24"
ORIGIN

  Query Match             8.0%;  Score 99.4;  DB 2;  Length 169500;
  Best Local Similarity   50.6%;  Pred. No. 3.7e-07;
  Matches  297;  Conservative    0;  Mismatches  281;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |          || ||   |   ||   |      |   | | |
Db      49667 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 49608

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
                | ||  |    | || || |||||  ||||  |||||  |   |||||  | |||  | |
Db      49607 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 49548

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |||||    | ||  || | |   ||   ||| || | ||||      || || |||
Db      49547 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 49494

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  ||||| |||   |       |||||  | |  ||||||| |||   ||| | | | |
Db      49493 GGCAAAGCCTTTAACCATTCCTCACACCTTGCTACACATAAGGTAATTCATACTGGAGAG 49434

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||||| | ||| |||  ||  |||||||||||   ||||          ||||  | |  |||
Db      49433 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 49374

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              ||| | |  |  |||||          |    || |||||| |||    |||| |||
Db      49373 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 49314

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              |  |      |  | |   ||||| | | | |   | |  ||| |||    |  |
Db      49313 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 49254

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||||  |||||| |  | |||         |||||| ||  |  ||||||||  |||
Db      49253 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 49194

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
              ||  |||||      ||   |    ||   |        |||||| ||  ||
Db      49193 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 49134

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
              ||||| |      ||  ||     |||  | ||    | | | || |
Db      49133 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTATAAGT 49087


RESULT 15
AK122869
LOCUS       AK122869               2237 bp    mRNA    linear   PRI 09-SEP-2003
DEFINITION  Homo sapiens cDNA FLJ16502 fis, clone FEBRA2006664, moderately
            similar to Zinc finger protein 43.
ACCESSION   AK122869
VERSION     AK122869.1  GI:34528067
KEYWORDS    oligo capping; fis (full insert sequence).
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M.,
            Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K.,
            Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S.,
            Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T.,
            Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M.,
            Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
            Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y.,
            Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.
  TITLE     NEDO human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 2237)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUL-2003) Takao Isogai, FLJ Project(HRI Team); 2-6-7
            Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: HRI and
            RAB; annotation: HRI and RAB.
FEATURES              Location/Qualifiers
     source           1. .2237
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="FEBRA2006664"
                      /tissue_type="brain"
                      /clone_lib="FEBRA2"
                      /dev_stage="fetal"
                      /note="cloning vector: pME18SFL3"
ORIGIN
```

Query Match 7.9%; Score 98.2; DB 9; Length 2237;
Best Local Similarity 52.0%; Pred. No. 1.6e-06;
Matches 273; Conservative 0; Mismatches 243; Indels 9; Gaps 2;

```
Qy        69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
             | || |    | || || |||||  ||||  |||||  |   |||||  | |||  | |
Db      1652 AGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTTAC 1711

Qy       129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
               |||||    | ||  || | |    ||   ||| || | ||||       || || ||||
Db      1712 TGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGTGG 1765

Qy       189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
             |||||  |||  |      |||||  | |  |||||||  |||   ||| | | | | ||
Db      1766 CAAAGCCTTTAACCATTCCTCACACCTTGCTACACATAAGGTAATTCATACTGGAGAGAA 1825

Qy       249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
             |||  |  |||  |||   ||  ||||||||||    ||||         ||||  | |  |||||
Db      1826 ACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGACA 1885

Qy       309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
             | | |  |  |||||              |    || |||||| |||    ||||  ||| |
Db      1886 TAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTTAA 1945

Qy       369 TAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
               |       |   | |    |||| | | | |   |  |  ||| |||   |  |
Db      1946 CCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAACC 2005

Qy       426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
             ||||  ||||||  |  | |||        |||||| ||  |  |||||||  |||  |
Db      2006 TAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAATTA 2065

Qy       486 TCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAAC 545
             |  |||||      ||    |       ||   |      |||||| ||  || ||
Db      2066 TGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCACAC 2125

Qy       546 TTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
             ||| |      ||  ||     |||  | ||    | | | || |
Db      2126 TTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTATAAGT 2170
```

Search completed: May 9, 2004, 08:56:51
Job time : 5094.16 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 03:47:06 ; Search time 545.661 Seconds
(without alignments)
9646.138 Million cell updates/sec

Title: US-09-831-804-2
Perfect score: 1239
Sequence: 1 atgagtgaaagtgacgaaac.....catcagtgatttctcgataa 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1239	100.0	2060	3	AAA15398	Aaa15398 DNA encod
2	1235.8	99.7	1239	6	ABZ32206	Abz32206 Candida a
3	99.4	8.0	2132	7	ADA53158	Ada53158 Human cod
4	96.8	7.8	1705	8	ACA98970	Aca98970 cDNA enco
5	96.6	7.8	2320	7	ACA56456	Aca56456 Human sig
6	95.2	7.7	4563	5	AAS91317	Aas91317 DNA encod
7	94.6	7.6	2597	7	AAL51569	Aal51569 Human nuc
8	92.4	7.5	3639	5	AAS64586	Aas64586 DNA encod
9	91.8	7.4	2110	9	ADB63624	Adb63624 Human cDN
10	91.8	7.4	2760	9	ADC56695	Adc56695 Human mac
11	91.8	7.4	3078	9	ADB62468	Adb62468 Human cDN
12	90.4	7.3	2026	7	AAD55863	Aad55863 Human nuc
13	90.4	7.3	2114	7	ADA53124	Ada53124 Human cod
14	90.2	7.3	2230	7	ADA52931	Ada52931 Human cod
15	89.2	7.2	2064	9	ADC30762	Adc30762 Human nov
16	89.2	7.2	2622	5	AAS68872	Aas68872 DNA encod
17	89.2	7.2	2729	4	AAH16178	Aah16178 Human cDN
18	89.2	7.2	3839	6	ABK83826	Abk83826 Human cDN
19	89	7.2	2298	4	AAH16608	Aah16608 Human cDN
20	89	7.2	2905	5	AAS66143	Aas66143 DNA encod
21	89	7.2	3020	5	AAS92560	Aas92560 DNA encod
22	89	7.2	3502	4	AAI57845	Aai57845 Human pol
23	88.6	7.2	2597	7	AAD55855	Aad55855 Human nuc
24	88.6	7.2	4227	8	ACA98938	Aca98938 cDNA enco
25	88.4	7.1	1952	9	ADC58104	Adc58104 Zinc fing
26	88.4	7.1	2509	6	AAD31103	Aad31103 Human tra
27	87.6	7.1	1757	6	ABQ81135	Abq81135 TRAF6-inh
28	87.2	7.0	831	6	ABQ55095	Abq55095 Human ova
29	87	7.0	976	5	AAS69188	Aas69188 DNA encod
30	86.8	7.0	2662	6	ABQ93353	Abq93353 Human cDN
31	86.8	7.0	6219	7	ACC46324	Acc46324 Human dit
32	86.8	7.0	6316	7	ACC46373	Acc46373 Human dit
33	86.2	7.0	1377	7	ABX34443	Abx34443 Human mdd
34	86.2	7.0	2558	4	AAS26690	Aas26690 Human gen
c 35	86.2	7.0	2558	4	AAS26691	Aas26691 Human gen
36	86.2	7.0	2558	7	ABX74039	Abx74039 Human nov
c 37	86.2	7.0	2558	7	ABX74040	Abx74040 Human nov
38	86.2	7.0	2607	5	AAS87125	Aas87125 DNA encod
39	86.2	7.0	2681	7	ABX34772	Abx34772 Human mdd
40	86	6.9	1890	9	ADC30336	Adc30336 Human nov
41	86	6.9	3309	7	ACC46347	Acc46347 Human dit
42	86	6.9	3400	9	ADB62883	Adb62883 Human cDN
43	85.6	6.9	1549	5	AAS67562	Aas67562 DNA encod
44	85.6	6.9	2476	7	ACC46407	Acc46407 Human dit
45	84.8	6.8	2239	7	ADA53516	Ada53516 Human cod

ALIGNMENTS

RESULT 1
AAA15398
ID AAA15398 standard; DNA; 2060 BP.
XX
AC AAA15398;
XX
DT 04-SEP-2000 (first entry)
XX
DE DNA encoding a transcription factor designated CATFIIIA.
XX
KW Transcription factor; CATFIIIA; DNA-binding protein;
KW ribosomal RNA 5S gene; fungal infection; ss.
XX
OS Candida albicans.
XX
FH Key Location/Qualifiers
FT CDS 720. .1958
FT /*tag= a
FT /transl_except= (pos: 1296. .1298, aa: Ser)
FT /transl_except= (pos: 1734. .1736, aa: Ser)
XX
PN WO200028037-A1.
XX
PD 18-MAY-2000.
XX
PF 09-NOV-1999; 99WO-FR002739.
XX
PR 10-NOV-1998; 98FR-00014147.
XX
PA (HMRI) HOECHST MARION ROUSSEL.
XX
PI Bordon-Pallier F, Camier S, Sentenac A;
XX
DR WPI; 2000-376549/32.
DR P-PSDB; AAY93316.
XX
PT New nucleic acid encoding Candida albicans transcription factor, useful
PT e.g. in screening for antimycotic agents and for immunization.
XX
PS Claim 4; Page 32-33; 45pp; French.
XX
CC The present sequence encodes a Candida albicans transcription factor,
CC designated CATFIIIA. The polypeptide is a DNA-binding protein, which is
CC involved in initiating transcription of the ribosomal RNA 5S gene. The
CC polynucleotide is used to screen for its specific inhibitors, potentially
CC useful as antimycotic agents, to raise an antibody response that is

CC protective against fungal infection and to raise antibodies. Such
CC antibodies, as well as the polypeptides and polynucleotides are used in
CC compositions for diagnosing and treating fungal infections, e.g. by
CC detecting polymorphisms and mutations
XX
SQ Sequence 2060 BP; 726 A; 315 C; 354 G; 665 T; 0 U; 0 Other;

Query Match 100.0%; Score 1239; DB 3; Length 2060;
Best Local Similarity 100.0%; Pred. No. 2.5e-234;
Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        720 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 779

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        780 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 839

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        840 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 899

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        900 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 959

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        960 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 1019

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1020 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 1079

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1080 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 1139

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1140 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 1199

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1200 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 1259

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1260 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 1319

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1320 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 1379

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1380 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 1439

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1440 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 1499

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1500 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 1559

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1560 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 1619

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1620 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 1679

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1680 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1739

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1740 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1799

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1800 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1859

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1860 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1919

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239
              ||||||||||||||||||||||||||||||||||||||
Db       1920 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1958
```

RESULT 2
ABZ32206
ID ABZ32206 standard; DNA; 1239 BP.
XX
AC ABZ32206;
XX
DT 30-JAN-2003 (first entry)
XX
DE Candida albicans essential gene SEQ ID NO 6493.
XX
KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal; gene; ss.
XX
OS Candida albicans.
XX
PN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US049486.
XX
PR 29-DEC-2000; 2000US-0259128P.
PR 20-FEB-2001; 2001US-00792024.
PR 22-AUG-2001; 2001US-0314050P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX
DR WPI; 2002-566694/60.
DR P-PSDB; ABP73656.
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele of
PT a gene and placing other allele of the gene under conditional expression.
XX
PS Claim 37; SEQ ID NO 6493; 167pp + Sequence Listing; English.
XX
CC The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene

CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans gene used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office
XX
SQ Sequence 1239 BP; 481 A; 171 C; 215 G; 372 T; 0 U; 0 Other;

Query Match 99.7%; Score 1235.8; DB 6; Length 1239;
Best Local Similarity 99.8%; Pred. No. 9.8e-234;
Matches 1237; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCA 60

Qy         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120

Qy        121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TTATTGGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAG 180

Qy        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240

Qy        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300

Qy        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360

Qy        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCATTTTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480

Qy        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 AAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTC 540

Qy        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCT 600

Qy        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 AAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGAT 660

Qy        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GATTCTACCATGATCAAAATATGGACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAG 720

Qy        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAAATGAATTAGTTGAACATTATAATATCTTCCATGATGGTAATATCCCTGATGATTTA 780

Qy        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TTAAAGGAAACTGAAGTGAAAAAATTAGAGAACCTATTAGATCAAGGATCGAAATTAAAT 840

Qy        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATTTGCATGAATTAGAAACAGAGAAATTAAAAGTGGAAGAAGATGAAGAAGATGAAGAA 900

Qy        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 GATAGTCTAGATGAAAAAAGAAGTGATGTTAGATCAGACTCAATGTCAGCTCAAAGATCA 960

Qy        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ATAAAATCATTTACTGCTTCTTTGGAAGGTTCAAAGAGTGTTTCTAAACTTATTCTGAAT 1020

Qy       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGTGGGAAGAAGATCAATTGTCCTAAGAATAATTGTGATAGAATGTTTTCTAGAGAATAT 1080

Qy       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GATTTACGTCGACATTTGAAATGGCATGATGATAATTTACAAAGAATTGAGTCATTCTTA 1140

Qy       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 AATAGTATAGAAAAAGAAGAAACTCCAGAAGGTGAACCATTGGTTAAAAAAGCCAGGATG 1200

Qy       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239
              |||||||||||||||||||||||||||||||||||||||
Db       1201 GATTTATTGCCAAATGAAACATCAGTGATTTCTCGATAA 1239
```

RESULT 3
ADA53158
ID ADA53158 standard; cDNA; 2132 BP.
XX
AC ADA53158;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human coding sequence, SEQ ID 726.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS Homo sapiens.
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-395539/38.
DR P-PSDB; ADA54797.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 726; 205pp; English.

XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2132 BP; 833 A; 423 C; 347 G; 529 T; 0 U; 0 Other;

Query Match 8.0%; Score 99.4; DB 7; Length 2132;
Best Local Similarity 50.6%; Pred. No. 3.3e-10;
Matches 297; Conservative 0; Mismatches 281; Indels 9; Gaps 2;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || |           || || |  | | || |   |     |  | | ||
Db       1384 GAATGTGGCAAAGCTTTTAACCAGTCCTCAATCCTTACTACACATAAGAGAATTCATACT 1443

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              | || |    | || ||| |||||  ||||  |||||  |    |||||  | |||  | ||
Db       1444 GGAGAGAAATCCTACAAATGTGAAGAATGTGGCAAAGCTTTCTATCGATCCTCAAAACTT 1503

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |||||   | ||  || | |   ||  ||| || | ||||      || || |||
Db       1504 ACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATGT------GAAGAATGT 1557

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              |  |||||  |||  |     |||||  | |  |||||||  |||   |||  | | | ||
Db       1558 GGCAAAGCCTTTAACCACTCCTCACACCTTGCTACACATAAGGTAATTCATACTGGAGAG 1617

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              |||||  | |||  |||   ||  ||||||||||   ||||          ||||  | |  |||
Db       1618 AAACCCTACCAATGTGAAGAATGTGGTAAAGCCTTTAACCAGTCCTCACACCTTACTAGA 1677

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||  |  |   |  |||||          |    || ||||||  |||    ||||  |||
Db       1678 CATAAGAGAATTCATACTGGAGAGAAACCCTACCAATGTGAAAAATGTGGCAAAGCTTTT 1737

Qy        367 TATAAACATCAATCTTTAA---GACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              | |   |     |  | |  |||||  | | | |   |  |  |||  |||    |  |
Db       1738 AACCAGTCCTCAAACCTTACTGGACATAAGAAAATTCATACTGGTGAGAAACTCTACAAA 1797

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||||  |||||| |  | |||          |||||||  ||  |  ||||||||  |||
Db       1798 CCTAAAAGATGTAACAGTGATTTTGAAAACACTTCAAAGTTTTCTAAACATAAAAGAAAT 1857

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
              ||  |||||      ||      |     ||   |          |||||||  ||  ||
Db       1858 TATGCTGGTGAGAAATCTTAGAAATGTGAAGAATGTAACAAAACCTTTAAAAGTTGTCAC 1917

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACT 590
              |||||  |      ||  ||     |||  | ||    | | | || |
Db       1918 ACTTGATTGTGCATAAGATAATTCATACTAAAAAAAAACCTATAAGT 1964
```

RESULT 4
ACA98970
ID ACA98970 standard; cDNA; 1705 BP.
XX
AC ACA98970;
XX
DT 25-JUL-2003 (first entry)
XX
DE cDNA encoding human nucleic acid-associated protein (NAAP) #51.
XX
KW Human; nucleic acid-associated protein; cytostatic; antiarteriosclerotic;
KW anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;
KW antiallergic; antiinflammatory; thyromimetic; gene therapy;
KW cell proliferative disorder; cancer; atherosclerosis;
KW neurological disorder; epilepsy; Huntington's disease; stroke;
KW immune disorder; inflammatory disorder; AIDS; allergy;
KW developmental disorder; Hypothyroidism; Cushing's syndrome; infection;
KW protein-protein interaction; drug-target interaction;
KW gene expression profile; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003023003-A2.
XX
PD 20-MAR-2003.
XX
PF 05-SEP-2002; 2002WO-US028540.
XX
PR 07-SEP-2001; 2001US-0317792P.
PR 07-SEP-2001; 2001US-0317912P.
PR 14-SEP-2001; 2001US-0322270P.
PR 21-SEP-2001; 2001US-0324040P.
PR 28-SEP-2001; 2001US-0326732P.
PR 19-OCT-2001; 2001US-0346716P.
PR 25-JAN-2002; 2002US-0351749P.
PR 22-FEB-2002; 2002US-0359498P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Tang YT, Jackson JL, Griffin JA, Elliott VS, Forsythe IJ;
PI Becha SD, Richardson TW, Lee EA, Sprague WW, Emerling BM;
PI Thangavelu K, Warren BA, Tran UK, Yue H, Xu Y, Yue H, Li JX;
PI Hafalia AJA, Sanjanwala B, Marquis JP, Gorvad AE, Lee SY, Ison CH;
PI Baughn MR, Chawla NK, Nguyen DB, Swarnakar A, Zebarjadian Y, Shah P;
PI Thornton M, Yao MG, Khan FA, Gandhi AR, Yang J, Kable AE;
PI Burford N, Ramkumar J;
XX
DR WPI; 2003-313243/30.
DR P-PSDB; ABU96722.
XX
PT New human nucleic acid associated proteins (NAAP), useful for diagnosing,
PT treating and preventing diseases or conditions associated with the
PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
PT infections.
XX
PS Claim 5; Page 340-341; 345pp; English.
XX
CC The invention describes a novel human isolated nucleic acid-associated
CC polypeptide (NAAP). The polypeptides and polynucleotides are useful in
CC diagnosing, treating and preventing diseases or conditions associated
CC with the decreased expression or overexpression of NAAP, such as cell
CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
CC disorders, or infections. These are also useful in assessing the effects
CC of exogenous compounds on the expression of nucleic acid and amino acid
CC sequences of NAAP. The NAAP or its fragments are useful in screening
CC compounds for effectiveness as agonist or antagonist of the polypeptides,
CC or in altering the expression of the target polynucleotide and compounds
CC that specifically bind to or modulate the activity of the polypeptide.
CC The microarray is useful in monitoring or measuring protein-protein
CC interactions, drug-target interactions, and gene expression profiles.
CC This sequence encodes a novel human nucleic acid-associated protein
CC (NAAP)
XX
SQ Sequence 1705 BP; 622 A; 305 C; 314 G; 464 T; 0 U; 0 Other;

Query Match 7.8%; Score 96.8; DB 8; Length 1705;
Best Local Similarity 52.9%; Pred. No. 1.1e-09;
Matches 259; Conservative 0; Mismatches 222; Indels 9; Gaps 2;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||  | ||||     |  ||     ||     || ||      |     |  |  |
Db        695 GAATGTCGCAAAACATTTTGCATGCTTTCACACCTAACTCAACATAAAAGAATCCAAACT 754

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              |  | || | | ||  || |||||  | ||  | |||||  |||   |  | |||    |
Db        755 AGAGTGAATTTCTACAAATGTGAAGCATATGGAAGAGCCTTTAACTGGTCCTCAACCCTT 814

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
              |  |||||    ||||  || | |    ||    ||| || |||||||  |       || |||
Db        815 AATAAACATAAGAGAATTCATACTGGAGAAAAACCTTACAAATGTAAA------GAATGT 868
```

```
Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
Db        869 GGCAAAGCCTTTAACCAGACCTCACACCTTATTAGACATAAGAGAATTCATACTGAAGAG 928

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
Db        929 AAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTCATCGACCCTTACTACA 988

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
Db        989 CATAATATAATTCATACTGGGGAAATTCCCTACAAATGTGAGAAATGTGTTAGAGCTTTT 1048

Qy        367 TATAAACATCAATCTTTAAGACATCATATATTATCTGTT---CATGAAAAAACATTAACG 423
Db       1049 AACCAAGCCTCAAAGCTTACTGAACATAAGTTAATTCATACCGGAGAGAAATGCTACAAA 1108

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
Db       1109 TGTGAAGAATGTGGCAAAGTTTTTAACCGATCTTCAAACCTTACAGACCATAAAAAAATT 1168

Qy        484 CATCATGGTG 493
Db       1169 CATACAGGAG 1178
```

RESULT 5
ACA56456
ID ACA56456 standard; cDNA; 2320 BP.
XX
AC ACA56456;
XX
DT 06-JUN-2003 (first entry)
XX
DE Human signalling pathway polynucleotide probe SEQ ID NO 1054.
XX
KW Human; probe; ss; array element; Parkinson's disease;
KW signalling pathway population; cancer; adenocarcinoma; leukaemia;
KW immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.
XX
OS Homo sapiens.
XX
PN US6500938-B1.
XX
PD 31-DEC-2002.
XX
PF 30-JAN-1998; 98US-00016434.
XX
PR 30-JAN-1998; 98US-00016434.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Au-Young J, Seilhamer JJ;
XX
DR WPI; 2003-352189/33.
XX
PT Combination of polynucleotide probes, useful as array elements in a
PT microarray for monitoring the expression of a number of target
PT polynucleotides.
XX
PS Claim 1; SEQ ID NO 1054; 65pp; English.
XX
CC The invention relates to a combination which, comprises a number of
CC polynucleotide probes comprising a sequence selected from one of the 1490
CC sequences mentioned in the specification. The combination is useful as an
CC array element in a microarray for monitoring the expression of a number
CC of target polynucleotides. The microarray is particularly useful in the
CC diagnosis and treatment of cancer and immunopathology and neuropathology.
CC The microarray is useful in diagnostics and treatment regimens, drug
CC discovery and development, toxicological and carcinogenicity studies,
CC forensics and pharmacogenomics. The microarray is also useful for
CC monitoring progression of diseases and for developing sophisticated
CC profiles for the effects of currently available therapeutic drugs. The
CC combination is also useful for purifying a subpopulation of mRNAs, cDNAs
CC and genomic fragments and in research and diagnostic applications. The
CC array can detect changes in expression in a large number of genes coding
CC for different signaling pathway populations which can be used to diagnose
CC various diseases including cancer e.g. adenocarcinoma and leukaemia,
CC immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
CC and Parkinson's disease. The present sequence represents a polynucleotide
CC probe of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=06500938B1
XX
SQ Sequence 2320 BP; 903 A; 423 C; 412 G; 582 T; 0 U; 0 Other;

```
  Query Match             7.8%;  Score 96.6;  DB 7;  Length 2320;
  Best Local Similarity   51.3%;  Pred. No. 1.2e-09;
  Matches  306;  Conservative    0;  Mismatches  279;  Indels   12;  Gaps    3;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
Db        964 AGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTTAC 1023

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
Db       1024 TACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGTGG 1077

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
Db       1078 CAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAGAA 1137

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
Db       1138 ACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACACA 1197

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
Db       1198 TGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTTAA 1257

Qy        369 TAAACA---TCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
Db       1258 TCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAATG 1317

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
Db       1318 TAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATTCA 1377

Qy        486 TCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAAC 545
Db       1378 TACTGGAGAGAAGCCTTACAAATCTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCAAA 1437

Qy        546 TTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATG 605
Db       1438 ACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAATG 1494

Qy        606 TGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
Db       1495 TGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGA 1551
```

RESULT 6
AAS91317
ID AAS91317 standard; cDNA; 4563 BP.
XX
AC AAS91317;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #27121.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.

XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR P-PSDB; ABG27130.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 1; SEQ ID NO 27121; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC coding sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4563 BP; 1752 A; 814 C; 824 G; 1173 T; 0 U; 0 Other;

Query Match 7.7%; Score 95.2; DB 5; Length 4563;
Best Local Similarity 52.7%; Pred. No. 2.5e-09;
Matches 258; Conservative 0; Mismatches 223; Indels 9; Gaps 2;

Qy 7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
Db 1315 GAATGTGGCAAAGCCTTTAACCAGTCCTCACACCTTACTCAACATAAAACAATTCATACT 1374
Qy 67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
Db 1375 GGAGGGAAAACCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCATCTTTCAGCCCTT 1434
Qy 127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
Db 1435 ACTAAACATAAGATAATTCATACTGGGGAGAAACCATACAAATGT------GAAGAATGT 1488
Qy 187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
Db 1489 GGCAAAGCTTTTAGGCAATCCTCACACCTTACTAGACATAAAGCAATTCATACTGGAGAG 1548
Qy 247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
Db 1549 AAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCATTTCTCAGACCTTAGAAGA 1608
Qy 307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
Db 1609 CATAAGATAATTCATACTGGAAAGAAACCCTACAAATGTGAAGAATGTGGGAAAGCTTTT 1668
Qy 367 TATAAACATCAATCTTTAAGA---CATCATATATTATCTGTTCATGAAAAAACATTAACG 423
Db 1669 AGCCAGTCCTCAACCCTTAGAAACCATCAGATAATTCATACTGGAGAGAAACCCTACAAA 1728
Qy 424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
Db 1729 TGTGAAGAATGTGGTAAAGCTTTTAAGTGGTCATCAAAACTTACTGTACATAAGGTAATT 1788
Qy 484 CATCATGGTG 493
Db 1789 CATACTGGAG 1798

RESULT 7
AAL51569
ID AAL51569 standard; DNA; 2597 BP.
XX
AC AAL51569;
XX
DT 10-APR-2003 (first entry)
XX
DE Human nucleic acid-associated protein coding sequence - SEQ ID No 52.
XX
KW Human; gene; ds; nucleic acid-associated protein; NAAP; arteriosclerosis;
KW cell proliferative disorder; atherosclerosis; cirrhosis; hepatitis; AIDS;
KW cancer; developmental disorder; renal tubular acidosis; anaemia; asthma;
KW mental retardation; neurological disorder; Alzheimer's disease; epilepsy;
KW Parkinson's disease; autoimmune disorder; inflammatory disorder; allergy;
KW Crohn's disease; transgenic animal; animal model.
XX
OS Homo sapiens.
XX
PN WO2003000864-A2.
XX
PD 03-JAN-2003.
XX
PF 20-JUN-2002; 2002WO-US021179.
XX
PR 22-JUN-2001; 2001US-0300518P.
PR 29-JUN-2001; 2001US-0301787P.
PR 29-JUN-2001; 2001US-0301792P.
PR 29-JUN-2001; 2001US-0301892P.
PR 29-JUN-2001; 2001US-0301893P.
PR 06-JUL-2001; 2001US-0303405P.
PR 06-JUL-2001; 2001US-0303442P.
PR 15-MAR-2002; 2002US-0364438P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Gandhi AR, Swarnakar A, Hafalia AJA, Warren BA, Emerling BM;
PI Arvizu CS, Ison CH, Honchell CD, Lee EA, Yue H, Forsythe IJ;
PI Ramkumar J, Griffin JA, Yang J, Sanjanwala MM, Baughn MR;
PI Borowsky ML, Yao MG, Walia NK, Bandman O, Lal PG, Becha SD, Lee SY;
PI Richardson TW, Elliott VS, Luo W, Tang YT, Zebarjadian Y, Lu Y;
XX
DR WPI; 2003-201420/19.
DR P-PSDB; AAO16419.
XX
PT New nucleic acid-associated proteins and polynucleotides, useful for
PT diagnosing, treating or preventing cell proliferative (e.g. cancer),
PT neurological (e.g. epilepsy or Parkinson's disease), or autoimmune
PT disorders (e.g. AIDS).
XX
PS Claim 12; Page 292-293; 312pp; English.
XX
CC The invention comprises the amino acid and coding sequences of human
CC nucleic acid-associated proteins (NAAP). The DNA and protein sequences of
CC the invention are useful for diagnosing, treating or preventing disorders
CC associated with aberrant expression of NAAP, such as: cell proliferative
CC disorders (e.g. arteriosclerosis, atherosclerosis, cirrhosis, hepatitis

```
CC   or cancer); developmental disorders (e.g. renal tubular acidosis, anaemia
CC   or mental retardation); neurological disorders (e.g. Alzheimer's disease,
CC   Parkinson's disease or epilepsy); and autoimmune/inflammatory disorders
CC   (e.g. AIDS, allergies, asthma or Crohn's disease). The DNA sequences of
CC   the invention are useful for creating transgenic animals to model human
CC   disease. The present DNA sequence encodes a human nucleic acid-associated
CC   protein of the invention
XX
SQ   Sequence 2597 BP; 947 A; 504 C; 505 G; 641 T; 0 U; 0 Other;

  Query Match             7.6%;  Score 94.6;  DB 7;  Length 2597;
  Best Local Similarity   55.1%;  Pred. No. 3e-09;
  Matches  231;  Conservative    0;  Mismatches  179;  Indels    9;  Gaps    2;

Qy         78 TTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTT 137
Db       1070 TTATAAATGTGAAGAATGTGGCAAAGCCTTTAATGTGTCCTCAACCCTTACTCAACATAA 1129

Qy        138 AAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATT 197
Db       1130 GAGAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAGTGTGGCAAAGCCTT 1183

Qy        198 TTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCA 257
Db       1184 TAACGTGTCCTCAACTCTTACTCAACATAAGAGAATTCATACTGGAGAAAAACCATACAA 1243

Qy        258 TTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCAC 317
Db       1244 ATGTGAAGAATGTGGCAAAGCCTTTAACACATCCTCACACCTCACCACACATAAAAGAAT 1303

Qy        318 CCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAA---ACA 374
Db       1304 TCATACCGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTTAACCAGTTCTC 1363

Qy        375 TCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATG 434
Db       1364 ACAACTTACTACACATAAGATAATTCATACTGGAGAGAAACCCTACAAATGTAAAGAATG 1423

Qy        435 TAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTG 493
Db       1424 TGGCAAAGCTTTTAAGCGGTCCTCAAACCTTACTGAACATAGGATAATTCATACTGGAG 1482


RESULT 8
AAS64586
ID   AAS64586 standard; cDNA; 3639 BP.
XX
AC   AAS64586;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #390.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG00399.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 1; SEQ ID NO 390; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human diagnostic
CC   coding sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3639 BP; 1403 A; 708 C; 622 G; 906 T; 0 U; 0 Other;

  Query Match             7.5%;  Score 92.4;  DB 5;  Length 3639;
  Best Local Similarity   54.1%;  Pred. No. 8.6e-09;
  Matches  236;  Conservative    0;  Mismatches  191;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
Db       738 AGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCATTCTTCAACCCTTGC 797

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
Db       798 TAAACATAAGAGAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 851

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
Db       852 CAAAGCTTTTAGCCGTTCTTCAACCCTTGCTAAACATAAGAGAATTCATACTGGAGAGAA 911

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
Db       912 ACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAGCAATTCCTCAACCCTTGCTAATCA 971

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
Db       972 TAAGATAACTCATACTGAAGAGAAACCCTACAAATGTAAAGAATGTGACAAAGCTTTTAA 1031

Qy        369 TAAACATCAATC---TTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
Db       1032 GCGACTCTCAACCCTTACTAAACATAAAATAATACATGCTGGAGAGAAACTCTACAAATG 1091

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
Db       1092 TGAAGAATGTGGCAAAGCTTTTAATCGATCTTCAAATCTTACTATACATAAGTTTATTCA 1151

Qy        486 TCATGGTGGATCTCCT 501
Db       1152 TACTGGAGAGAAACCT 1167


RESULT 9
ADB63624
ID   ADB63624 standard; cDNA; 2110 BP.
XX
```

```
AC   ADB63624;
XX
DT   04-DEC-2003  (first entry)
XX
DE   Human cDNA encoding clone THYMU20071120.
XX
KW   Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW   tissue regeneration; cell regeneration; membrane protein;
KW   signal transduction-related protein; transcription-related protein;
KW   osteoporosis; neurological disease; cancer; tumour.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             170. .1792
FT                   /*tag=  a
FT                   /product= "Clone THYMU20071120 protein"
XX
PN   EP1308459-A2.
XX
PD   07-MAY-2003.
XX
PF   28-MAR-2002; 2002EP-00007401.
XX
PR   05-NOV-2001; 2001JP-00379298.
PR   25-JAN-2002; 2002US-00350978.
XX
PA   (HELI-) HELIX RES INST.
PA   (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI   Isogai T,  Sugiyama T,  Otsuki T,  Wakamatsu A,  Sato H,  Ishii S;
PI   Yamamoto J,  Isono Y,  Hio Y,  Otsuka K,  Nagai K,  Irie R,  Tamechika I;
PI   Seki N,  Yoshikawa T,  Otsuka M,  Nagahari K,  Masuho Y;
XX
DR   WPI; 2003-450961/43.
DR   P-PSDB; ADB65594.
XX
PT   New polynucleotides and polypeptides, useful for developing a diagnostic
PT   marker or medicines for regulation of their expression and activity, or
PT   as targets of gene therapy.
XX
PS   Claim 1; Page; 222pp; English.
XX
CC   The invention discloses a polynucleotide comprising a sequence selected
CC   from 1970 fully defined nucleotide sequences which encode novel
CC   polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC   or its partial peptide, an antibody binding to the polypeptide or peptide
CC   of the polynucleotide, immunologically assaying the polypeptide or
CC   peptide of the polynucleotide by contacting the polypeptide or peptide
CC   with the antibody of the encoded protein, and observing the binding
CC   between the two, a transformant carrying the polynucleotide in an
CC   expressible manner and an antisense polynucleotide. The oligonucleotide
CC   is useful as a primer for synthesising the polynucleotide, or as a probe
CC   for detecting the polynucleotide. The polynucleotides and encoded
CC   proteins are useful as pharmaceutical agents and many disease-related
CC   genes may be included in them, for developing a diagnostic marker or
CC   medicines for regulation of their expression and activity, or as targets
CC   of gene therapy. The genes are involved in tissue and/or cell
CC   regeneration. Membrane proteins, signal transduction-related proteins,
CC   transcription-related proteins, disease-related proteins and genes
CC   encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC   neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC   the activity or expression of the encoded protein to treat diseases. The
CC   sequence presented is a cDNA of the invention. Note: Some of the sequence
CC   data for this patent is not represented in the printed specification, but
CC   is based on sequence information supplied by the European Patent Office.
XX
SQ   Sequence 2110 BP; 784 A; 394 C; 386 G; 546 T; 0 U; 0 Other;

  Query Match             7.4%;  Score 91.8;  DB 9;  Length 2110;
  Best Local Similarity   51.7%;  Pred. No. 1e-08;
  Matches  314;  Conservative    0;  Mismatches  277;  Indels   16;  Gaps    4;

Qy        61 CGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCA 120
Db      1253 CATACCGAAGAGAAACCCTACAAATGTAAAGAATGTGGCAAAGCTTTTAAACACTCCTCA 1312

Qy       121 TTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGAC 180
Db      1313 GCCCTTACTACACATAAGAGAATTCACACTGGAGAGAAACCCTACAAATGT------GAA 1366

Qy       181 GATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCC 240
Db      1367 GAATGTGGCAAAGCTTTTAACCGATCCTCAAAACTTACTGAACATAAGAAACTTCATACT 1426

Qy       241 GAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTG 300
Db      1427 GGAAAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTATCCAATCCTCAAAACTT 1486

Qy       301 AAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAA 360
Db      1487 ACTGAACATAAAAAAATTCATTCTGGAGAGATACCCTACAAGTGTGAAGAATGTGGCAAA 1546

Qy       361 GCATTTTATAAACATCAATCTTTA---AGACATCATATATTATCTGTTCATGAAAAAACA 417
Db      1547 GCTTTTAAACACTCCTCATCCCTTACTACACATAAAAGAATTCATACTGGGGAGAAACCC 1606

Qy       418 TTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAA 477
Db      1607 TACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCGATCCTCAAAACTTACTGAACATAAG 1666

Qy       478 TTAAAACATCATGGTG-GATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAA 536
Db      1667 ATAATTCATACTGGAGAGAAACCCTATAAATGTGAGAGATGTGACAAAGCTTTTAACCAA 1726

Qy       537 TTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATG 596
Db      1727 TCTGCAAACCT------TACTAAACATAAAAAAATACATACTGGAGAGAAACTACAGAAC 1780

Qy       597 TCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCA 656
Db      1781 TGGAATGTGTGATAATGATTTTGACAACACCTCAAATTTTTCTAAATATAAAGGAAATCA 1840

Qy       657 TGATGAT 663
Db      1841 TACTGGT 1847


RESULT 10
ADC56695
ID   ADC56695 standard; cDNA; 2760 BP.
XX
AC   ADC56695;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human macroprotein-45-76 cDNA.
XX
KW   human; macroprotein-45.76; dementia; arrhythmia; asthma; diabetes; ss;
KW   gene.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             180. .1430
FT                   /*tag=  a
FT                   /product= "Human macroprotein-45.76"
XX
PN   CN1382725-A.
XX
PD   04-DEC-2002.
XX
PF   26-APR-2001; 2001CN-00112764.
XX
PR   26-APR-2001; 2001CN-00112764.
XX
```

PA (BIOW-) BIOWINDOW GENE DEV INC SHANGHAI.
XX
PI Mao Y, Xie Y;
XX
DR WPI; 2003-269486/27.
DR P-PSDB; ADC56696.
XX
PT New human macroprotein-45.76, encoding polynucleotide, antagonist and
PT recombinant production, useful for treating dementia, arrhythmia, asthma
PT and diabetes.
XX
PS Claim 6; SEQ ID NO 1; 34pp; Chinese.
XX
CC The invention relates to a novel human macroprotein-45.76, the encoding
CC polynucleotide, an antagonist and a method of recombinant production. The
CC protein of the invention may be useful for treating dementia, arrhythmia,
CC asthma and diabetes. The current sequence is that of the human
CC macroprotein-45.76 cDNA of the invention.
XX
SQ Sequence 2760 BP; 1016 A; 463 C; 499 G; 782 T; 0 U; 0 Other;

Query Match 7.4%; Score 91.8; DB 9; Length 2760;
Best Local Similarity 52.1%; Pred. No. 1.1e-08;
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;

```
Qy          11 GTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCCAAAA 70
               ||| | || |    |  ||    ||   | || ||     |     |  | | | | |
Db         712 GTGGCAAATCATTTTGCATGCTTTCACAATTAACTCAACATAAGAAAATTCATACTAGAG 771

Qy          71 AGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGC 130
               |||||  || || || |||||  |||| ||||||||| |||   |  | |||      |
Db         772 AGTATTCTTACAAATGTGAAGAATGTGGTAAAGCCTTTAACTGGTCCTCAACCCTTACTA 831

Qy         131 AACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATA 190
               |||||    | ||  || | |    ||   ||| || ||||||       || || ||||  |
Db         832 AACATAAGATAATTCATACTGGAGAAAAACCCTACAAATGT------GAAGAATGTGGCA 885

Qy         191 AAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAC 250
               |||| |||  | |    ||| || |    | |||||   ||   ||| | | | | ||||
Db         886 AAGCTTTTAACCGGTCCTCAAATCTTACTAAACATAAAATAATTCATACTGGAGAGAAAC 945

Qy         251 CATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATG 310
               | | | | |||  ||  ||||| ||||   ||||          | | | |  | ||||
Db         946 CCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCGGTCCTCAACCCTTACTAAACATA 1005

Qy         311 AAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATA 370
               |||  |  |||||| |          | || |||| | |||     | || ||| |
Db        1006 AAAGAATTCATACAGAAGAGAAACCCTACAAATGTGAAGAATGTGGCAAGGCCTTTAACC 1065

Qy         371 AACATCAATCTTT---AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTA 427
               |         | |    | |||| | | | |   | |  | || ||| | |   |  |||
Db        1066 AGTTCTCGATTCTTAATAAACATAAGAGAATTCATATGGAAGATAAACCCTACAAATGTG 1125

Qy         428 AACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATC 487
               || |||||   ||||  || |    |   ||||   |   | |||||||  |||  |||
Db        1126 AAGAATGTGGCAAAGCCTTTAGAGTATTCTCAATTCTTAAAAAACATAAGATAATCCATA 1185

Qy         488 ATGGTGGATCTCC 500
               ||| | |   ||
Db        1186 CTGGGGAAAAACC 1198
```

RESULT 11
ADB62468
ID ADB62468 standard; cDNA; 3078 BP.
XX
AC ADB62468;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human cDNA encoding clone FEBRA20063720.
XX
KW Human; ss; gene; pharmaceutical; diagnostic; gene therapy;
KW tissue regeneration; cell regeneration; membrane protein;
KW signal transduction-related protein; transcription-related protein;
KW osteoporosis; neurological disease; cancer; tumour.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 118. .1878
FT /*tag= a
FT /product= "Clone FEBRA20063720 protein"
XX
PN EP1308459-A2.
XX
PD 07-MAY-2003.
XX
PF 28-MAR-2002; 2002EP-00007401.
XX
PR 05-NOV-2001; 2001JP-00379298.
PR 25-JAN-2002; 2002US-00350978.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-450961/43.
DR P-PSDB; ADB64438.
XX
PT New polynucleotides and polypeptides, useful for developing a diagnostic
PT marker or medicines for regulation of their expression and activity, or
PT as targets of gene therapy.
XX
PS Claim 1; Page; 222pp; English.
XX
CC The invention discloses a polynucleotide comprising a sequence selected
CC from 1970 fully defined nucleotide sequences which encode novel
CC polypeptides. Also claimed is a polypeptide encoded by the polynucleotide
CC or its partial peptide, an antibody binding to the polypeptide or peptide
CC of the polynucleotide, immunologically assaying the polypeptide or
CC peptide of the polynucleotide by contacting the polypeptide or peptide
CC with the antibody of the encoded protein, and observing the binding
CC between the two, a transformant carrying the polynucleotide in an
CC expressible manner and an antisense polynucleotide. The oligonucleotide
CC is useful as a primer for synthesising the polynucleotide, or as a probe
CC for detecting the polynucleotide. The polynucleotides and encoded
CC proteins are useful as pharmaceutical agents and many disease-related
CC genes may be included in them, for developing a diagnostic marker or
CC medicines for regulation of their expression and activity, or as targets
CC of gene therapy. The genes are involved in tissue and/or cell
CC regeneration. Membrane proteins, signal transduction-related proteins,
CC transcription-related proteins, disease-related proteins and genes
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,
CC neurological diseases, cancer, tumours. The cDNA may be used to regulate
CC the activity or expression of the encoded protein to treat diseases. The
CC sequence presented is a cDNA of the invention. Note: Some of the sequence
CC data for this patent is not represented in the printed specification, but
CC is based on sequence information supplied by the European Patent Office.
XX
SQ Sequence 3078 BP; 1159 A; 516 C; 538 G; 865 T; 0 U; 0 Other;

Query Match 7.4%; Score 91.8; DB 9; Length 3078;
Best Local Similarity 52.1%; Pred. No. 1.1e-08;
Matches 257; Conservative 0; Mismatches 227; Indels 9; Gaps 2;

```
Qy          11 GTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCCAAAA 70
               ||| | || |    |  ||    ||   | || ||     |     |  | | | | |
Db         650 GTGGCAAATCATTTTGCATGCTTTCACAATTAACTCAACATAAGAAAATTCATACTAGAG 709

Qy          71 AGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGC 130
               |||||  || || || |||||  |||| ||||||||| |||   |  | |||      |
Db         710 AGTATTCTTACAAATGTGAAGAATGTGGTAAAGCCTTTAACTGGTCCTCAACCCTTACTA 769
```

```
Qy        131 AACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATA 190
              |||||    | ||  || | |   ||   ||| || ||||||       || || ||||  |
Db        770 AACATAAGATAATTCATACTGGAGAAAAACCCTACAAATGT------GAAGAATGTGGCA 823

Qy        191 AAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAC 250
              ||||  |||  | |    ||| || |    |||||   ||   ||| | | | ||||
Db        824 AAGCTTTTAACCGGTCCTCAAATCTTACTAAACATAAAATAATTCATACTGGAGAGAAAC 883

Qy        251 CATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATG 310
              | | | | |||  ||  ||||| ||||   ||||          | | | |  | ||||
Db        884 CCTACAAATGTGAAGAATGTGGCAAAGCTTTTAACCGGTCCTCAACCCTTACTAAACATA 943

Qy        311 AAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATA 370
              |||  |  |||||| |           |  | || |||| | |||    | ||| ||| |
Db        944 AAAGAATTCATACAGAAGAGAAACCCTACAAATGTGAAGAATGTGGCAAGGCCTTTAACC 1003

Qy        371 AACATCAATCTTT---AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTA 427
              |      |  |   | |||| | | | |  | | || ||| | |  |  |||
Db       1004 AGTTCTCGATTCTTAATAAACATAAGAGAATTCATATGGAAGATAAACCCTACAAATGTG 1063

Qy        428 AACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATC 487
              || |||||  ||||  || |   |  ||||  |  | |||||||  |||  |||
Db       1064 AAGAATGTGGCAAAGCCTTTAGAGTATTCTCAATTCTTAAAAAACATAAGATAATCCATA 1123

Qy        488 ATGGTGGATCTCC 500
               ||| | |  ||
Db       1124 CTGGGGAAAAACC 1136
```

RESULT 12
AAD55863
ID AAD55863 standard; cDNA; 2026 BP.
XX
AC AAD55863;
XX
DT 07-AUG-2003 (first entry)
XX
DE Human nucleic acid associated protein (NAAP)-34 cDNA.
XX
KW Human; nucleic acid associated protein; NAAP; stroke; AIDS; nootropic;
KW cancer; atherosclerosis; neurological; epilepsy; Huntington's disease;
KW developmental disorder; antiinflammatory; neuroprotective; thyromimetic;
KW Cushing's syndrome; infection; gene therapy; cytostatic; anticonvulsant;
KW cerebroprotective; hypothyroidism; cell proliferative disorder; allergy;
KW gene; ss.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   126. .1973
FT   /*tag= a
FT   /product= "Human NAAP protein"
XX
PN WO2003006618-A2.
XX
PD 23-JAN-2003.
XX
PF 10-JUL-2002; 2002WO-US021971.
XX
PR 12-JUL-2001; 2001US-0305089P.
PR 12-JUL-2001; 2001US-0305104P.
PR 13-JUL-2001; 2001US-0305325P.
PR 13-JUL-2001; 2001US-0305390P.
PR 19-JUL-2001; 2001US-0306960P.
PR 20-JUL-2001; 2001US-0306694P.
PR 27-JUL-2001; 2001US-0308170P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Swarnakar A, Richardson TW, Warren BA, Griffin JA, Tang YT;
PI Yue H, Baughn MR, Emerling BM, Lal PG, Lu DAM, Forsythe IJ;
PI Ramkumar J, Li JX, Becha SD, Duggan BM, Sanjanwala MM, Lee EA;
PI Burford N, Elliott VS, Ison CH, Ding L, Borowsky ML, Yao MG;
PI Barroso I, Tran B, Walia NK, Hafalia AJA, Nguyen DB, Lu Y;
PI Arvizu CS;
XX
DR WPI; 2003-221732/21.
DR P-PSDB; AAE37047.
XX
PT New human nucleic acid associated proteins (NAAP), useful for diagnosing,
PT treating and preventing diseases or conditions associated with the
PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
PT infections.
XX
PS Claim 5; Page 259-260; 260pp; English.
XX
CC The invention relates to human nucleic acid associated proteins (NAAP)
CC and their corresponding nucleic acid sequences. The invention is useful
CC in diagnosing, treating and preventing diseases or conditions associated
CC with the decreased expression or overexpression of NAAP, such as cell
CC proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
CC disorders, or infections. It is also useful in assessing the effects of
CC exogenous compounds on the expression of nucleic acid and amino acid
CC sequences of NAAP. The NAAP or its fragments are useful in screening
CC compounds for which acts as their agonist or antagonist. The microarray
CC is useful in monitoring or measuring protein-protein interactions, drug-
CC target interactions, and gene expression profiles. NAAP DNA is used in
CC gene therapy. The present sequence is human NAAP cDNA
XX
SQ Sequence 2026 BP; 699 A; 422 C; 396 G; 509 T; 0 U; 0 Other;

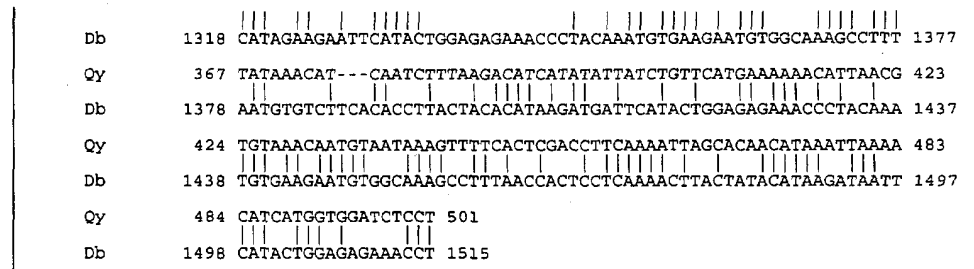
Query Match 7.3%; Score 90.4; DB 7; Length 2026;
Best Local Similarity 51.8%; Pred. No. 2e-08;
Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||  |     || ||     | || |   |     |  | | | |
Db       1002 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 1061

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              | || |   | || || |||||  |||| ||||||  | ||| |   || |||     |
Db       1062 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 1121

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
                |||||      |  ||   ||| ||   ||| || ||||||      || || |||
Db       1122 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 1175

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
              || ||||| |||  | ||   ||| |  |     | |||||   ||   |||||  | | |
Db       1176 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 1235

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
              ||| | | | | |||   |  |||||  |||||  ||||  |     |  | | | || |
Db       1236 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 1295

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
              |||  ||  |  |||||                |  | || |||| | |||    |||| |||
Db       1296 CATAGAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 1355

Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
              ||      |  ||   |  | |||| | ||  |   |  |   || ||| | |  |
Db       1356 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 1415

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
              ||| || |||||   ||||  || |  |  | ||||||  |  |   ||||||  |||
Db       1416 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 1475

Qy        484 CATCATGGTGGATCTCCT 501
              |||  ||| |    |||
Db       1476 CATACTGGAGAGAAACCT 1493
```

RESULT 13

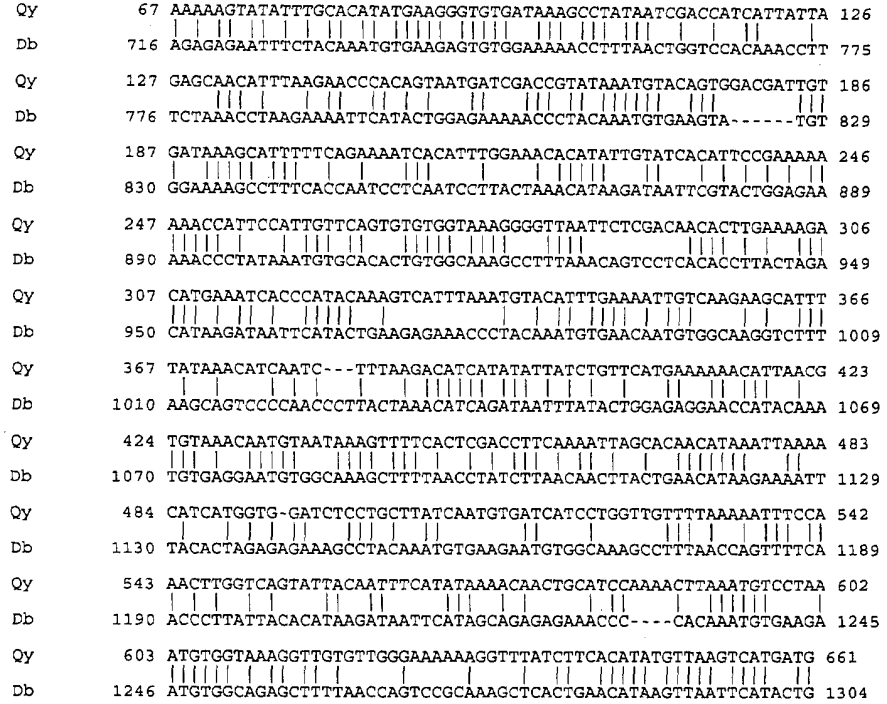
ADA53124
ID ADA53124 standard; cDNA; 2114 BP.
XX
AC ADA53124;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human coding sequence, SEQ ID 692.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS Homo sapiens.
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-395539/38.
DR P-PSDB; ADA54763.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 692; 205pp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2114 BP; 717 A; 439 C; 434 G; 524 T; 0 U; 0 Other;

Query Match 7.3%; Score 90.4; DB 7; Length 2114;
Best Local Similarity 51.8%; Pred. No. 2e-08;
Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;



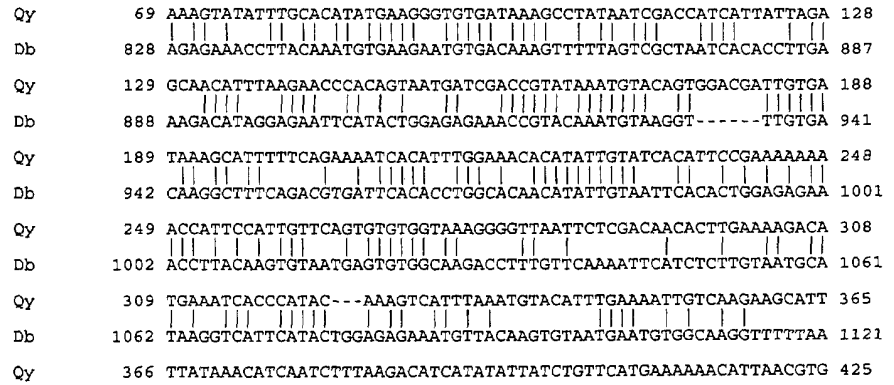
RESULT 14
ADA52931
ID ADA52931 standard; cDNA; 2230 BP.
XX
AC ADA52931;
XX
DT 20-NOV-2003 (first entry)
XX
DE Human coding sequence, SEQ ID 499.
XX
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW inflammatory disease; osteoporosis; neurological disease; gene; ss.
XX
OS Homo sapiens.
XX
PN EP1293569-A2.
XX
PD 19-MAR-2003.
XX
PF 21-MAR-2002; 2002EP-00006586.
XX
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
PA (HELI-) HELIX RES INST.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
XX
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I;
PI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
XX
DR WPI; 2003-395539/38.
DR P-PSDB; ADA54570.
XX
PT New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 499; 205pp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2230 BP; 831 A; 434 C; 400 G; 565 T; 0 U; 0 Other;

Query Match 7.3%; Score 90.2; DB 7; Length 2230;
Best Local Similarity 50.4%; Pred. No. 2.2e-08;
Matches 332; Conservative 0; Mismatches 313; Indels 14; Gaps 4;



RESULT 15
ADC30762
ID ADC30762 standard; cDNA; 2064 BP.
XX
AC ADC30762;
XX
DT 18-DEC-2003 (first entry)
XX
DE Human novel cDNA sequence, SEQ ID NO:844.
XX
KW Human; diagnostic; drug screening; forensics; gene mapping;
KW biodiversity assessment; Parkinson's disease; Alzheimer's disease;
KW neurodegenerative diseases; anaemia; platelet disorder; wound; burns;
KW ulcers; osteoporosis; autoimmune disease; cancer;
KW molecular weight marker; food supplement; antiparkinsonian; nootropic;
KW neuroprotective; antianaemic; anticoagulant; thrombolytic; vulnerary;
KW antiulcer; osteopathic; immunosuppressive; antiinflammatory; cytostatic;
KW gene therapy; chromosome 19q12; gene; ss.
XX
OS Homo sapiens.
XX
PN WO2003029271-A2.
XX
PD 10-APR-2003.
XX
PF 24-SEP-2002; 2002WO-US030474.
XX
PR 24-SEP-2001; 2001US-0324631P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang TY, Zhang J, Ren F, Xue AJ, Zhao QA, Wang J, Wehrman T;
PI Zhou P, Ghosh M, Wang D, Ma Y, Asundi V, Wang Z, Weng G;
PI Haley-Vicente D, Drmanac RT;
XX
DR WPI; 2003-371981/35.
DR P-PSDB; ADC31733.
XX
PT New polynucleotide and polypeptide useful for diagnosing, preventing or
PT treating conditions such as neurodegenerative diseases, anemias, platelet
PT disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
PT cancer.
XX
PS Claim 1; SEQ ID NO 844; 1185pp; English.
XX
CC The invention relates to 971 novel human cDNA sequences (ADC29919-
CC ADC30889)and the polypeptides they encode (ADC30890-ADC31860). The
CC invention also relates to nucleic acid sequences over 99% identical with
CC the novel human cDNAs. The invention additionally encompasses expression
CC vectors and host cells comprising a nucleic acid of the invention; the
CC recombinant production of a polypeptide of the invention; an antibody
CC against a polypeptide of the invention; a method of detecting
CC polynucleotides or polypeptides of the invention; and methods of
CC identifying a compound which binds to a polypeptide of the invention. The
CC invention further discloses methods of peventing, treating or
CC ameliorating a medical condition; kits comprising polynucleotide probes
CC and/or monoclonal antibodies for carrying out the methods of the
CC invention; methods for the identification of compounds that modulate the
CC expression or activity of the polynucleotide and/or polypeptide; and 767
CC contig sequences corresponding to the cDNA sequences of the invention
CC (ADC31861-ADC32627) and the polypeptides encoded by the contigs (ADC32628
CC -ADC33394). The nucleic acids and polypeptides of the invention are
CC useful in diagnostics, drug screening, forensics, gene mapping, in the
CC identification of mutations responsible for genetic disorders or other
CC traits, for assessing biodiversity, and in producing many other types of
CC data and products dependent on DNA and amino acid sequences. They are
CC also used for treating diseases such as Parkinson's disease, Alzheimer's
CC disease and other neurodegenerative diseases, anaemia, platelet
CC disorders, wounds, burns, ulcers, osteoporosis, autoimmune diseases or
CC cancer. The nucleic acids may also be used as hybridisation probes or
CC primers, and in the recombinant production of a protein. The polypeptides
CC are also useful in generating antibodies, as molecular weight markers,
CC and as food supplements. The present sequence represents a specifically
CC claimed human cDNA sequence of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2064 BP; 703 A; 375 C; 449 G; 537 T; 0 U; 0 Other;

Query Match 7.2%; Score 89.2; DB 9; Length 2064;
Best Local Similarity 53.7%; Pred. No. 3.4e-08;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;



```
          | | |||    |   |  | | |||||||| | |   |  |   || ||| | |  | |||
Db       1122 TCACAAATCAAACCTTGCATGTCATCATAGACTTCATACTGGAGAGAAACCTTACAAGTG 1181

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
          |||  |||||   || ||||| | | ||   |||||  | | ||| ||| |   |   ||
Db       1182 TAATGAATGTGGCAAGGTTTTTAATTGAAAATCAAACCTTGAACATCATCATAGAGTTCA 1241

Qy        486 TCATGGTGGATCTCCT 501
          |   || | |   |||
Db       1242 TATCGGAGAAAAACCT 1257
```

Search completed: May 9, 2004, 06:07:07
Job time : 549.661 secs

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: May 9, 2004, 05:31:45 ; Search time 3619.38 Seconds
(without alignments)
10222.531 Million cell updates/sec

Title: US-09-831-804-2
Perfect score: 1239
Sequence: 1 atgagtgaaagtgacgaaac.....catcagtgatttctcgataa 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	182.2	14.7	628	28	BZ298000	BZ298000 CG3957.f1
c	2	167.6	13.5	959	29	CNS06MMP	AL405671 T7 end of
	3	98.2	7.9	1609	11	AK032220	AK032220 Mus muscu
	4	97.8	7.9	2647	11	BC047646	BC047646 Homo sapi
	5	97.6	7.9	745	9	AU123448	AU123448 AU123448
	6	92.4	7.5	2672	11	BC036394	BC036394 Homo sapi
	7	92	7.4	591	28	AQ005136	AQ005136 CIT-HSP-2
	8	91	7.3	593	28	B99387	B99387 CIT-HSP-228
	9	91	7.3	1163	13	BX414627	BX414627 BX414627
	10	90.4	7.3	2694	11	BC022527	BC022527 Homo sapi
	11	90.4	7.3	2698	11	BC037782	BC037782 Homo sapi
	12	90.2	7.3	2253	11	BC028252	BC028252 Mus muscu
	13	89.2	7.2	604	10	BE161630	BE161630 MR3-HT044
	14	89.2	7.2	3609	11	BC032590	BC032590 Homo sapi
	15	88.6	7.2	535	9	AL705393	AL705393 DKFZp686M
	16	88.6	7.2	2174	11	AK033001	AK033001 Mus muscu
	17	88.6	7.2	2330	11	BC037426	BC037426 Homo sapi
	18	88.6	7.2	2622	11	AK033958	AK033958 Mus muscu
	19	88.4	7.1	1084	13	BX456765	BX456765 BX456765
	20	88	7.1	895	13	BU195416	BU195416 AGENCOURT
	21	87.6	7.1	832	13	BX437291	BX437291 BX437291
	22	87.6	7.1	1201	13	BX355654	BX355654 BX355654
	23	87.4	7.1	1200	13	BX437758	BX437758 BX437758
	24	87.2	7.0	579	28	AQ347265	AQ347265 RPCI11-11
c	25	87.2	7.0	687	28	AQ389266	AQ389266 RPCI11-15
	26	87.2	7.0	907	13	BQ423752	BQ423752 AGENCOURT
	27	87	7.0	660	14	CD770140	CD770140 AGENCOURT
	28	87	7.0	842	13	BU507408	BU507408 AGENCOURT
c	29	86.8	7.0	521	28	AQ475626	AQ475626 CITBI-E1-
	30	86.8	7.0	705	28	AQ194282	AQ194282 RPCI11-61
c	31	86.8	7.0	860	28	AQ749175	AQ749175 HS_5575_A
	32	86.8	7.0	897	13	BQ431141	BQ431141 AGENCOURT
	33	86.8	7.0	1906	11	AK013043	AK013043 Mus muscu
	34	86.6	7.0	694	12	BG696505	BG696505 602658829
	35	86.4	7.0	874	13	BU171453	BU171453 AGENCOURT
	36	86.4	7.0	2791	11	BC020045	BC020045 Homo sapi
c	37	86.2	7.0	523	14	CA874049	CA874049 K0929E01-
c	38	86	6.9	557	10	BF817669	BF817669 MR2-CI012
	39	86	6.9	667	14	CF135920	CF135920 UI-HF-BN0
c	40	86	6.9	685	29	AG149704	AG149704 Pan trogl
	41	86	6.9	801	14	CD657347	CD657347 AGENCOURT
	42	85.6	6.9	711	14	CB154257	CB154257 K-EST0212
	43	85.6	6.9	712	9	AI792344	AI792344 an30d02.y
	44	85.4	6.9	682	12	BG818171	BG818171 602779744
	45	85.4	6.9	2106	11	AK030776	AK030776 Mus muscu

ALIGNMENTS

RESULT 1
BZ298000/c
LOCUS BZ298000 628 bp DNA linear GSS 31-OCT-2002
DEFINITION CG3957.f1 Candida glabrata Random Genomic Library Candida glabrata genomic clone CG3957, genomic survey sequence.
ACCESSION BZ298000
VERSION BZ298000.1 GI:24440936
KEYWORDS GSS.
SOURCE Candida glabrata
ORGANISM Candida glabrata
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; mitosporic Saccharomycetales; Candida.
REFERENCE 1 (bases 1 to 628)
AUTHORS Wong,S., Fares,M.A., Zimmermann,W., Butler,G. and Wolfe,K.H.
TITLE Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast Candida glabrata
JOURNAL Genome Biol. 4 (2), R10 (2003)
MEDLINE 22508158
PUBMED 12620120
COMMENT Contact: Wong S
Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Fax: 353 1 6798558
Email: swong@tcd.ie
Class: plasmid ends.

```
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Candida glabrata"
                     /mol_type="genomic DNA"
                     /strain="CBS 138"
                     /db_xref="taxon:5478"
                     /clone="CG3957"
                     /clone_lib="Candida glabrata Random Genomic Library"
ORIGIN

  Query Match             14.7%;  Score 182.2;  DB 28;  Length 628;
  Best Local Similarity   59.5%;  Pred. No. 3.2e-24;
  Matches  348;  Conservative    0;  Mismatches  228;  Indels    9;  Gaps    2;

Qy          84 ATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGAAC 143
               ||| ||    ||||| || |||| ||    | || || ||  |     ||||     |  ||
Db         625 ATACGATAATTGTGACAAGGCCTTTACAAGGCCGTCCTTGCTCACAGAACACCAGAATAC 566

Qy         144 CCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAG 203
                 ||  | |   |  |    || | |     | | |||||  |   |||||    |
Db         565 ---AGTGCATCTTGGGAGGAAACCTTGGAAATGTAATCAATGTGAAAGTTCATTTACTAA 509

Qy         204 AAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGTTC 263
               |||    || || || | |||  |     ||||  | ||  ||| ||| ||  |||||||
Db         508 AAAGATCCACTTAGAGAGACACTTATACACACACAGATGAAAGACCGTTTTATTGTTC 449

Qy         264 AGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATAC 323
               | | ||||| || ||| | | | || | |||||  ||||| |||||||  | || || ||
Db         448 ATTTTGTGGGAAGGGGCTAATTACTAGGCAACAACTGAAACGACATGAGGTAACTCACAC 389

Qy         324 AAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTT 383
               || || || || ||| || |||| |||| | ||    || || || || | |      ||
Db         388 CAAATCGTTCAATTGTGAATATGAAGGTTGTAATGAGAGTTTCTACAAGCACCCTCAATT 329

Qy         384 AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGT 443
               |||   |||||| ||| | ||||||  | |||   ||| ||| |  |||| || |||
Db         328 AAGGGCTCATATTTTAGCAGTTCATTTACAAAGTCTAAAATGTCACGAATGCAACAAAAG 269

Qy         444 TTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACATCAT------GGTGGATC 497
               |||     ||||||  | | |  | | || | |   |||| |||      | ||
Db         268 TTTTCAAAGACCTTACAGACTCAAAAATCACATAGCCAAACACCATAATCCAGATGTTGT 209

Qy         498 TCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATT 557
               ||||||||||||||          || ||||| ||| | |||| |||| ||||| |   ||
Db         208 CAATGCTTATCAATGTACCTTCAGTGTTTGTTCTAAGAGTTTCAAAACATGGTCTGCGTT 149

Qy         558 ACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTG 617
               |  ||| ||| | ||| |    ||||| ||  | ||||| |||| |||  |||||   ||
Db         148 AAGATTACATGTTAAAAATGATCATCCGAAGTTAAAATGCCCTATATGCAGTAAACCCTG 89

Qy         618 TGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
               ||||||| || | |||||||  |   || |||  ||  ||||
Db          88 TGTTGGGGAAGACGGTTTAAATATGCACATGAAAATCCACGATGA 44


RESULT 2
CNS06MMP/c
LOCUS       CNS06MMP                 959 bp    DNA     linear   GSS 17-JUN-2001
DEFINITION  T7 end of clone AU0AA005F10 of library AU0AA from strain CBS 3082
            of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION   AL405671
VERSION     AL405671.1  GI:12168715
KEYWORDS    GSS.
SOURCE      Saccharomyces kluyveri
  ORGANISM  Saccharomyces kluyveri
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE   1  (bases 1 to 959)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 959)
  AUTHORS   Neuveglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F.,
            Gaillardin,C. and Casaregola,S.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 9.
            Saccharomyces kluyveri
  JOURNAL   FEBS Lett. 487 (1), 56-60 (2000)
  MEDLINE   20584719
   PUBMED   11152884
REFERENCE   3  (bases 1 to 959)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .959
                     /organism="Saccharomyces kluyveri"
                     /mol_type="genomic DNA"
                     /strain="CBS 3082"
                     /db_xref="taxon:4934"
                     /clone="AU0AA005F10"
                     /clone_lib="AU0AA"
                     /note="end : T7"
     misc_feature    complement(<71. .>958)
                     /note="similar to Saccharomyces cerevisiae ORF YPR186c [
                     PZF1 ; TFIIIA (transcription initiation factor) ]"
                     /evidence=not_experimental
ORIGIN

  Query Match             13.5%;  Score 167.6;  DB 29;  Length 959;
  Best Local Similarity   57.2%;  Pred. No. 1.6e-21;
  Matches  324;  Conservative    1;  Mismatches  235;  Indels    6;  Gaps    1;

Qy         211 CATTTGGAAACACATATTGTATCACATTCCGAAAAAAAAACCATTCCATTGTTCAGTGTGT 270
               || ||||||| ||||||   |  | |||||||| || |||||||| ||||||:|    |||
Db         958 CACTTGGAAAGACATATGTTCACGCATTCCGATAATAAACCATTTCATTGTYCCTATTGT 899

Qy         271 GGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACATGAAATCACCCATACAAAGTCA 330
               |||||||| |||| | |    || || || ||||||||||| ||||| ||||| || ||
Db         898 GGTAAAGGTGTTACTACAAAGCAGCAGTTAAAAAGACATGAGATCACACATACCAAATCG 839

Qy         331 TTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTATAAACATCAATCTTTAAGACAT 390
               || ||||||  || |||    ||  | |||   ||||| || |||       ||||
Db         838 TTCAAATGTGAATATGAGGGCTGCAACGAATGTTTTTACAAGCATTCTCAACTAAGGTCA 779

Qy         391 CATATATTATCTGTTCATGAAAAAACATTAACGTGTAAACAATGTAATAAAGTTTTCACT 450
               |||| | | || || ||      | | ||| || ||| | || ||   |||| | |||
Db         778 CATACACTTTCGGTACACTTGCAGAAATTGACCTGTGAGCATTGCGGTAAAATGTTCCAA 719

Qy         451 CGACCTTCAAAATTAGCACAACATAAATTAAAACATCATGGTGGATCTCCTGCT------ 504
               | || |     |   | | |||     ||| |||||   ||    |  ||
Db         718 AGGCCATATCGTCTGCAAAACCATCTTGCAAAGCATCACAATGTGGATGTTGAAAACAAA 659

Qy         505 TATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAAACTTGGTCAGTATTACAATTT 564
               |||||||||  | | | | ||    | | |||  |||| ||| | |   | ||
Db         658 TATCAATGTACTTACACGGCGTGCACAGAGACTTTTAAAACCTGGACTGCTCTTCAGCAG 599
```

```
Qy        565 CATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAATGTGGTAAAGGTTGTGTTGGG 624
              ||||| ||      |||||  ||  |  | |||      ||||||  ||||  |||||||||
Db        598 CATATCAAGGCGGACCATCCCAAGTTGCACTGTAACGTATGTGGCAAAGCTTGTGTTGGT 539

Qy        625 AAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGATTCTACCATGATCAAAATATGG 684
              |    |||  ||      || |||  |   |||||  ||  ||     | ||  ||||  |||
Db        538 GAGTCTGGTCTACAGATGCACATGCGAGTCCATGACGAGGCTTTAGTAATAAAAAATTGG 479

Qy        685 ACTTGTGATTATTGTGATGTGGGGAAATTTGCAAAGAAAAATGAATTAGTTGAACATTAT 744
              |  |||     ||||  |     |     ||||| |||||   |||  |  |    || ||
Db        478 AAATGTACCATTTGTCAGCAAGTCTCTTTTGCTAAGAAGGCTGATCTTTTATCTCACTAC 419

Qy        745 AATATCTTCCATGATGGTAATATCCC 770
              |        ||| | | | |||| ||
Db        418 ATGGAACATCATAAGGATGATATACC 393
```

RESULT 3
AK032220
LOCUS AK032220 1609 bp mRNA linear HTC 18-SEP-2003
DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length enriched library, clone:6430503O15 product:MSZF33 (FRAGMENT) homolog [Mus musculus], full insert sequence.
ACCESSION AK032220
VERSION AK032220.1 GI:26328050
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 1609)
AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/.
FEATURES Location/Qualifiers
source 1. .1609
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:6430503O15"
/db_xref="MGI:2395676"
/db_xref="taxon:10090"
/clone="6430503O15"
/sex="male"
/tissue_type="olfactory brain"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
CDS 330. .1223
/note="unnamed protein product; MSZF33 (FRAGMENT) homolog [Mus musculus] (SPTR|O88251, evidence: FASTY, 87.1%ID, 100%length, match=255)
putative"
/codon_start=1
/protein_id="BAC27766.1"
/db_xref="GI:26328051"
/translation="MLETYRNLNAVGYNWEDSNIEEHCESSRRHGRHERNHTGEKPYE
GIQYGEAFVHHSSLQMRKIIHTGEKRYKCNQCDKAYSRHSILQIHKRTHSGEKPYECN
QCGKAFTQHSHLKIHMVTHTGEKPYKCNQCGKAFACHNKLQKHERTHTGEKPYKCDQC
NKAFVYESYLQVHKKTHTGEKPYKCNECGKAFARHSHLKVHKITHTGEKPYKCNQCGK
ALAYHSTLQVHQRTHTGEKPYECEQCGKAFANQSYFQVHKRIHTGEKPYKCDQCGKAF
VGSSDLKRHERVHTGRETLQM"
polyA_signal 1594. .1599
/note="putative"
polyA_site 1609
/note="putative"
ORIGIN

Query Match 7.9%; Score 98.2; DB 11; Length 1609;
Best Local Similarity 51.2%; Pred. No. 1.5e-08;
Matches 310; Conservative 0; Mismatches 283; Indels 12; Gaps 3;

```
Qy         60 ACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATC 119
              || | |   | || |     | || ||    |    ||||||||||||||||    ||||
Db        521 ACATACTGGAGAGAAACGCTACAAATGCAATCAATGTGATAAAGCCTATTCACGACACAG 580

Qy        120 ATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGA 179
              |  || |   ||||  |||||| || |||    ||   ||| ||| ||||||      |
Db        581 CATTCTACAAATACATAAAAGAACACATAGTGGAGAGAAACCCTATGAATGTA------A 634

Qy        180 CGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTC 239
              | |||| ||||||  |||     | |    ||| |  ||| |||||  ||  |||||  |
Db        635 TCAATGTGGTAAAGCCTTTACACAACACAGTCATCTCAAAATACATATGGTTACACATAC 694
```

```
Qy        240 CGAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTT 299
Db        695 TGGAGAGAAACCTTACAAATGTAATCAATGTGGTAAAGCCTTTGCATGTCATAATAAACT 754

Qy        300 GAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGA 359
Db        755 TCAAAAACATGAAAGAACACATACTGGAGAGAAACCTTACAAGTGTGATCAATGTAATAA 814

Qy        360 AGCATTT---TATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAAC 416
Db        815 AGCCTTTGTATATGAAAGTTATTTACAAGTTCATAAAAAAACACATACTGGAGAGAAACC 874

Qy        417 ATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAA 476
Db        875 TTACAAATGTAATGAATGTGGTAAAGCCTTTGCACGACACAGTCATCTCAAAGTGCATAA 934

Qy        477 ATTAAAACATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAA 536
Db        935 AATAACACATACTGGAGAGAAAC---CTTACAAATGTAATCAATGTGGTAAAGCCCTTGC 991

Qy        537 TTTCCAAACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATG 596
Db        992 ATATCATAGTACACTCCAAGTACATCAAAGAACACATACTGGAGAGAAGCCCTATGAATG 1051

Qy        597 TCCTAAATGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCA 656
Db       1052 TGAGCAATGTGGCAAAGCCTTTGCAAATCAAAGTTATTTCCAAGTACATAAAAGAATACA 1111

Qy        657 TGATG 661
Db       1112 TACTG 1116
```

```
RESULT 4
BC047646
LOCUS       BC047646               2647 bp    mRNA    linear   HTC 19-NOV-2003
DEFINITION  Homo sapiens zinc finger protein 85 (HPF4, HTF1), mRNA (cDNA clone
            IMAGE:4828290), with apparent retained intron.
ACCESSION   BC047646
VERSION     BC047646.1  GI:29126804
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2647)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2647)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-MAR-2003) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 90 Row: i Column: 7
            This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source          1. .2647
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4828290"
                     /tissue_type="Testis"
                     /clone_lib="NIH_MGC_97"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
ORIGIN
```

```
  Query Match             7.9%;  Score 97.8;  DB 11;  Length 2647;
  Best Local Similarity   50.1%;  Pred. No. 1.7e-08;
  Matches  330;  Conservative    0;  Mismatches  317;  Indels   12;  Gaps    3;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
Db       1235 GAATGTGGCAAAACTTTTAACCGATTCTCAACTCTTACTACCCATAAGATAATTCATACT 1294

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
Db       1295 GGAGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTT 1354

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
Db       1355 ACTACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGT 1408

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
Db       1409 GGCAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAG 1468

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
Db       1469 AAACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACCACA 1528

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
Db       1529 CATGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTT 1588

Qy        367 TATAA---ACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
Db       1589 AATCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAA 1648

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
Db       1649 TGTAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATT 1708

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
Db       1709 CATACTGGAGAGAAGCCTTACAAATGTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCA 1768
```

```
Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAA 603
Db       1769 AAACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAA 1825

Qy        604 TGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCATGATGA 662
Db       1826 TGTGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGA 1884

RESULT 5
AU123448
LOCUS       AU123448                 745 bp    mRNA    linear   EST 01-AUG-2002
DEFINITION  AU123448 NT2RM2 Homo sapiens cDNA clone NT2RM2000319 5', mRNA
            sequence.
ACCESSION   AU123448
VERSION     AU123448.1  GI:10948164
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 745)
  AUTHORS   Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S., Saito,K., Yamamoto,J.,
            Nakamura,Y., Nishikawa,T., Nagai,T., Suzuki,Y., Sugano,S. and
            Isogai,T.
  TITLE     HRI human cDNA project (Ota,T., Wakamatsu,A., Ozawa,M., Ishii,S.,
            Saito,K., Yamamoto,J., Nakamura,Y., Nishikawa,T., Nagai,T.,
            Suzuki,Y., Sugano,S., Isogai,T.)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Takao Isogai
            Genomics Laboratory
            Helix Research Institute
            1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            Tel: 81-438-52-3975
            Fax: 81-438-52-3986
            Email: genomics@hri.co.jp
            HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
            Research Institute; cDNA library construction: Department of
            Virology, Institute of Medical Science, University of Tokyo, and
            Helix Research Institute.
FEATURES             Location/Qualifiers
     source          1..745
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="NT2RM2000319"
                     /cell_type="teratocarcinoma"
                     /cell_line="NT2"
                     /clone_lib="NT2RM2"
                     /note="Vector: pME18SFL3; mRNA from uninduced NT2 neuronal
                     precursor cells"
ORIGIN

  Query Match             7.9%;  Score 97.6;  DB 9;  Length 745;
  Best Local Similarity   50.2%;  Pred. No. 2.1e-08;
  Matches  328;  Conservative    0;  Mismatches  314;  Indels   12;  Gaps    3;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
Db         87 GAATGTGGCAAAACTTTTAACCGATCCTCAACTCTTACTACCCATAAGATAATTCATACT 146

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
Db        147 GGAGAGAAACCCTACAAATGTAAAGAATGTGGTAAAGCTTTTAACCGTTCTTCAACCCTT 206

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
Db        207 ACTACCCATAGAAAAATTCATACTGGAGAGAAACCTTACAAATGT------GAAGAATGT 260

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
Db        261 GGCAAAGCCTTTAAGCAGTCCTCAAACCTTACTACACATAAGATAATTCATACTGGAGAG 320

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
Db        321 AAACCCTACAAATGTAAAAAATGTGGAAAAGCCTTTAACCAGTCTGCACACCTTACTACA 380

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
Db        381 CATGAGGTAATTCATACTGGAGAGAAACCCTACAAATGTGAAAAATGTGGAAAAGCCTTT 440

Qy        367 TATAA---ACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
Db        441 AATCATTTCTCACACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCTTACAAA 500

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
Db        501 TGTAAAGAATGTGGTAAAGCTTTTAAACACTCTTCAACCCTTACTAAACATAAGATAATT 560

Qy        484 CATCATGGTGGATCTCCTGCTTATCAATGTGATCATCCTGGTTGTTTTAAAAATTTCCAA 543
Db        561 CATACTGGAGAGAAGCCTTACAAATGTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCA 620

Qy        544 ACTTGGTCAGTATTACAATTTCATATAAAACAACTGCATCCAAAACTTAAATGTCCTAAA 603
Db        621 AAACTTACTGAACATAAGAAAATTCATACTGGAGAGAAACCCTA---TGAATGTGAAAAA 677

Qy        604 TGTGGTAAAGGTTGTGTTGGGAAAAAAGGTTTATCTTCACATATGTTAAGTCAT 657
Db        678 TGTGGCAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCAT 731

RESULT 6
BC036394
LOCUS       BC036394                2672 bp    mRNA    linear   HTC 19-NOV-2003
DEFINITION  Homo sapiens zinc finger protein 85 (HPF4, HTF1), mRNA (cDNA clone
            IMAGE:5259399), containing frame-shift errors.
ACCESSION   BC036394
VERSION     BC036394.1  GI:23025784
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2672)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 2672)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-AUG-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
```

```
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
            cDNA Library Preparation: Michael J. Brownstein (NHGRI) &  Shiraki
            Toshiyuki and Piero Carninci (RIKEN)
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 73 Row: a Column: 16
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1..2672
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5259399"
                     /tissue_type="Brain, hippocampus"
                     /clone_lib="NIH_MGC_95"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript"
ORIGIN

  Query Match             7.5%;  Score 92.4;  DB 11;  Length 2672;
  Best Local Similarity   54.1%;  Pred. No. 1.7e-07;
  Matches  236;  Conservative    0;  Mismatches  191;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
Db       1745 AGAGAAGCCTTACAAATGTAAAGAATGTGAAAAAGCTTTTAACCAATCCTCAAAACTTAC 1804

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
Db       1805 TGAACATAAGAAAATTCATACTGGAGAGAAACCCTATGAATGT------GAAAAATGTGG 1858

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
Db       1859 CAAAGCTTTTAACCAGTCCTCAAATCTTACTAGACATAAGAAAAGTCATACAGAAGAGAA 1918

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
Db       1919 ACCTTACAAATGTGAAGAATGTGGCAAAGGTTTTAAATGGCCCTCAACCCTTACTATCCA 1978

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
Db       1979 TAAGATAATTCATACTGGAGAGAAACCATACAAATGTGAAGAATGTGGCAAAGCTTTTAA 2038

Qy        369 TAAA---CATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
Db       2039 CCAATCCTCAAAACTTACCAAACATAAGAAAATTCATACTGGAGAGAAACCCTACACATG 2098

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
Db       2099 TGAAGAATGTGGCAAAGCCTTTAACCAGTCCTCAAACCTTACTAAACATAAGAGAATTCA 2158

Qy        486 TCATGGTGGATCTCCT 501
Db       2159 TACTGGAGAAAAACCT 2174

RESULT 7
AQ005136
LOCUS       AQ005136                 591 bp    DNA     linear   GSS 27-JUN-1998
DEFINITION  CIT-HSP-2290O17.TF CIT-HSP Homo sapiens genomic clone 2290O17,
            genomic survey sequence.
ACCESSION   AQ005136
VERSION     AQ005136.1  GI:3082581
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Other_GSSs: CIT-HSP-2290O17.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21;
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..591
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7150844"
                     /db_xref="taxon:9606"
                     /clone="2290O17"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

  Query Match             7.4%;  Score 92;  DB 28;  Length 591;
  Best Local Similarity   52.2%;  Pred. No. 2.4e-07;
  Matches  256;  Conservative    0;  Mismatches  225;  Indels    9;  Gaps    2;

Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
Db         69 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 128

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
Db        129 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 188

Qy        127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
Db        189 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 242

Qy        187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
Db        243 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 302

Qy        247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
Db        303 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 362

Qy        307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
Db        363 CATAAAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 422

Qy        367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
Db        423 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 482

Qy        424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
Db        483 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 542
```

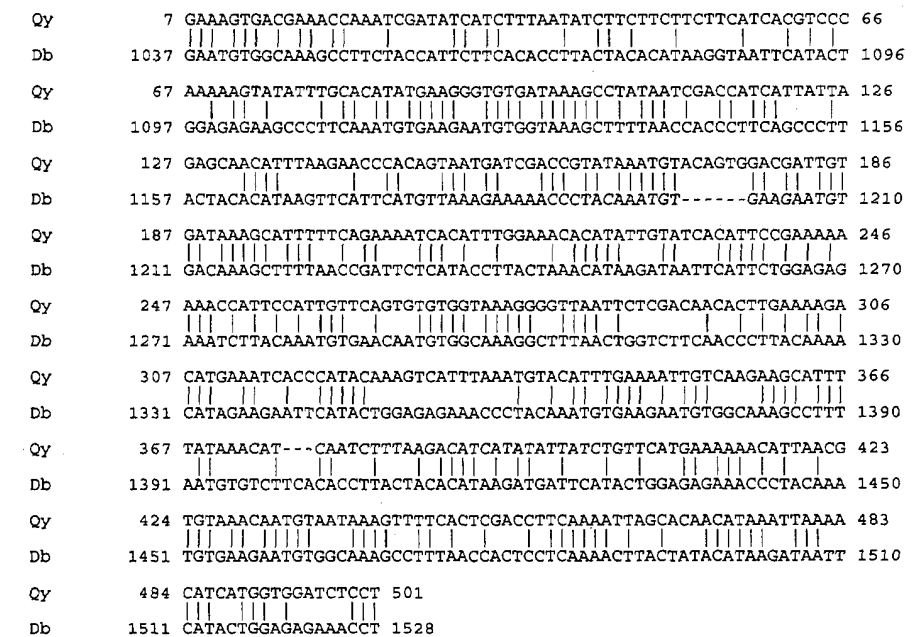
```
Qy        484 CATCATGGTG 493
              ||| ||| |
Db        543 CATACTGGAG 552


RESULT 8
B99387
LOCUS       B99387                   593 bp    DNA     linear   GSS 26-JUN-1998
DEFINITION  CIT-HSP-2281A10.TF CIT-HSP Homo sapiens genomic clone 2281A10,
            genomic survey sequence.
ACCESSION   B99387
VERSION     B99387.1  GI:3027197
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 593)
  AUTHORS   Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
            Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
            Simon,M. and Venter,J.C.
  TITLE     Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building (1998)
  JOURNAL   Unpublished (1998)
COMMENT     Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21;
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .593
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="GDB:7147044"
                     /db_xref="taxon:9606"
                     /clone="2281A10"
                     /sex="Male"
                     /cell_type="Sperm"
                     /clone_lib="CIT-HSP"
                     /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                     HindIII"
ORIGIN

  Query Match             7.3%;  Score 91;  DB 28;  Length 593;
  Best Local Similarity   54.4%;  Pred. No. 3.7e-07;
  Matches  231;  Conservative    0;  Mismatches  185;  Indels    9;  Gaps    2;

Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              |||| |    | || || |||||  ||||  ||||| | ||  |  | |||    |
Db         47 AAAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAGCCAGTCCTCAACCCTTAG 106

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
               |||||    | ||  || | |   ||   ||| || ||||||      || || ||||
Db        107 AAAACATGAGATAATTCATACTGGAGAGAAACCCTACAAATGT------GAAGAATGTGG 160

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              |||||| |||     |  |||||  |    | |||||  |||   ||| | ||| | ||
Db        161 TAAAGCTTTTAAGTGGTCCTCACACCTTACTAGACATAAAGTAATTCATACTGAAGAGAA 220

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||  ||  |||||  || |   ||||   |      |  | | |  |||
Db        221 ACCCTACAAATGTGAAGAATGTGGCAAGGCTTTTAACCATTTCTCAGCCCTTAGGAAACA 280

Qy        309 TGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTTTA 368
              | | || |  |||||             |  | || |||| | |||    ||||  |||

Db        281 TAAGATAATTCATACTGGAAAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCTTTTAG 340

Qy        369 TAAACATCAATCTTT---AAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
                |      | || |   || |||| | ||| |   |  |    || ||| | |  |  ||
Db        341 CCAGTCCTCAACTCTTAGAAAACATGAGATAATTCATACTGGAGAGAAACCCTACAAATG 400

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              | || |||||  ||||| ||| |   |  | ||||||  |  |    ||||||  |||  ||
Db        401 TGAAGAATGTGGTAAAGCTTTTAAGTGGTCCTCAAAACTTACTGTACATAAGGTAATTCA 460

Qy        486 TCATG 490
              |  ||
Db        461 TACTG 465


RESULT 9
BX414627
LOCUS       BX414627                 1163 bp    mRNA    linear   EST 13-MAY-2003
DEFINITION  BX414627 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP001YD12
            5-PRIME, mRNA sequence.
ACCESSION   BX414627
VERSION     BX414627.1  GI:30649903
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1163)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 9703.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0CAP001DB06QP1&cluster=9703.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0CAP001DB06QP1.
FEATURES             Location/Qualifiers
     source          1. .1163
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP001YD12"
                     /tissue_type="THYMUS"
                     /clone_lib="Homo sapiens THYMUS"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             7.3%;  Score 91;  DB 13;  Length 1163;
  Best Local Similarity   50.1%;  Pred. No. 3.5e-07;
  Matches  289;  Conservative    8;  Mismatches  268;  Indels   12;  Gaps    3;

Qy         89 AAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGAGCAACATTTAAGAACCCACA 148
              |||  |||| |||||| | ||| ||| | |||     |          |||  |:|::  || |
Db        121 AAGAATGTGGTAAAGCTTTTAACCGATCTTCAACCCTTACTACSCATAGARGRRTTCATA 180

Qy        149 GTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGATAAAGCATTTTTCAGAAAAT 208
              |   ||   ||| || ||||||      || || ||:|  |||||  |||         |
Db        181 CTGGAGAGAAACCTTACAAATGT------GAAGAATGKGGCAAAGCCTTTAAGCAGTCCT 234

Qy        209 CACATTTGGAAACACATATTGTATCACATTCCGAAAAAAAACCATTCCATTGTTCAGTGT 268
              || |  |     |||||||   ||   ||| | | | | ||||| | | | |||  | |
Db        235 CAAACCTTACTACACATAAGATAATTCATACTGGAGAGAAACCCTACAAATGTAAAAGAT 294
```

RESULT 10
BC022527
LOCUS BC022527 2694 bp mRNA linear HTC 19-NOV-2003
DEFINITION Homo sapiens hypothetical protein LOC148206, mRNA (cDNA clone IMAGE:4797857), containing frame-shift errors.
ACCESSION BC022527
VERSION BC022527.1 GI:18490307
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2694)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 2694)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (01-FEB-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: http://www-shgc.stanford.edu
Contact: (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers, R. M.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 32 Row: f Column: 2
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein
This clone has the following problem: frame shifted.
FEATURES Location/Qualifiers
source 1. .2694
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4797857"
/tissue_type="Brain, hypothalamus"
/clone_lib="NIH_MGC_96"
/lab_host="DH10B"
/note="Vector: pBluescript"
ORIGIN

Query Match 7.3%; Score 90.4; DB 11; Length 2694;
Best Local Similarity 51.8%; Pred. No. 4.1e-07;
Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;



RESULT 11
BC037782
LOCUS BC037782 2698 bp mRNA linear HTC 19-NOV-2003
DEFINITION Homo sapiens hypothetical protein LOC148206, mRNA (cDNA clone IMAGE:4797729), with apparent retained intron.
ACCESSION BC037782
VERSION BC037782.1 GI:23349076
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2698)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 2698)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre, BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 70 Row: g Column: 16
This clone was selected for full length sequencing because it passed the following selection criteria: Similarity but not identity to protein
This clone has the following problem: retained intron.
FEATURES Location/Qualifiers
source 1. .2698
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:4797729"
/tissue_type="Brain, hypothalamus"
/clone_lib="NIH_MGC_96"
/lab_host="DH10B"
/note="Vector: pBluescript"
ORIGIN

Query Match 7.3%; Score 90.4; DB 11; Length 2698;
Best Local Similarity 51.8%; Pred. No. 4.1e-07;
Matches 258; Conservative 0; Mismatches 231; Indels 9; Gaps 2;

```
Qy          7 GAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGTCCC 66
              ||| ||| | || ||   |     || ||     | || |   |     | | | |
Db       1042 GAATGTGGCAAAGCCTTCTACCATTCTTCACACCTTACTACACATAAGGTAATTCATACT 1101

Qy         67 AAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTA 126
              | || |   | || || |||||  |||| ||||||  | ||| |  || |||    |
Db       1102 GGAGAGAAGCCCTTCAAATGTGAAGAATGTGGTAAAGCTTTTAACCACCCTTCAGCCCTT 1161

Qy         127 GAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGT 186
               ||||      |  ||   ||| ||   ||| || ||||||      || || |||
Db       1162 ACTACACATAAGTTCATTCATGTTAAAGAAAAACCCTACAAATGT------GAAGAATGT 1215

Qy         187 GATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAA 246
               || ||||| ||| | ||   ||| | |   | ||||   ||   |||| | | |
Db       1216 GACAAAGCTTTTAACCGATTCTCATACCTTACTAAACATAAGATAATTCATTCTGGAGAG 1275

Qy         247 AAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGA 306
               ||| | | | | ||| |  |||| |||||  |||| |       | | | | || |
Db       1276 AAATCTTACAAATGTGAACAATGTGGCAAAGGCTTTAACTGGTCTTCAACCCTTACAAAA 1335

Qy         307 CATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATTT 366
               ||| ||  |  |||||           | | || |||| | |||     |||| |||
Db       1336 CATAGAAGAATTCATACTGGAGAGAAACCCTACAAATGTGAAGAATGTGGCAAAGCCTTT 1395

Qy         367 TATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACG 423
               ||     |   ||  |  | |||| | ||  |  |  |   || ||| | |  |
Db       1396 AATGTGTCTTCACACCTTACTACACATAAGATGATTCATACTGGAGAGAAACCCTACAAA 1455

Qy         424 TGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAA 483
               ||| || ||||||  ||||  || |  |   | |||||| | |  |  ||||||  |||
Db       1456 TGTGAAGAATGTGGCAAAGCCTTTAACCACTCCTCAAAACTTACTATACATAAGATAATT 1515

Qy         484 CATCATGGTGGATCTCCT 501
               |||  |||  |     |||
Db       1516 CATACTGGAGAGAAACCT 1533
```

RESULT 12
BC028252
LOCUS BC028252 2253 bp mRNA linear HTC 20-SEP-2002
DEFINITION Mus musculus, clone IMAGE:3674739, mRNA.
ACCESSION BC028252
VERSION BC028252.1 GI:20380246
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 2253)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (08-APR-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 65 Row: a Column: 17
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13386417
This clone has the following problem: frame shifted.

FEATURES Location/Qualifiers
source 1. .2253
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CZECH II"
/db_xref="taxon:10090"
/clone="IMAGE:3674739"
/tissue_type="Mammary tumor metastatized to lung. Tumor arose spontaneously from a senescent normal mammary (clonal) outgrowth infected with the virus MMTV."
/clone_lib="NCI_CGAP_Lu29"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"

ORIGIN

Query Match 7.3%; Score 90.2; DB 11; Length 2253;
Best Local Similarity 51.9%; Pred. No. 4.6e-07;
Matches 256; Conservative 0; Mismatches 228; Indels 9; Gaps 2;

```
Qy          4 AGTGAAAGTGACGAAACCAAATCGATATCATCTTTAATATCTTCTTCTTCTTCATCACGT 63
              | | || |||| || ||   || | |      ||     |      || |     | ||| |
Db        823 AATCAATGTGATAAAGCCTTTTCAACACACAGTTCTCTTCAAACTCATAGAAGAACACAT 882

Qy         64 CCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTA 123
                ||  || | |    | || || | |    |||||||||||||| |   ||   |   |
Db        883 ACCGGAAGGAAACCCTTCAAATGTAATCAATGTGATAAAGCCTTTTCCCGTAAACATTGT 942

Qy        124 TTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGAT 183
              |   |    ||| |||| |  || | |   ||   ||| |  |||||||      | |
Db        943 GTGCAAACCCATATAAGGATACATACTGGAGAAAAACCCTTCAAATGTA------ATCAA 996

Qy        184 TGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAA 243
              |||||||||||| ||||    | |      | |  |||| ||||    | ||||| ||| |
Db        997 TGTGATAAAGCCTTTTCACAAGACAGTACTCTTCAAACTCATAGAAGAACACATACCGGA 1056

Qy        244 AAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAA 303
              | ||||| ||| | |||     ||||| |||||   ||  |  |  |||     |  ||
Db       1057 GAGAAACCCTTCAAATGTAATCAGTGTGATAAAGCCTTTTCTGTTAAACATAGTCTCCAA 1116

Qy        304 AGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCA 363
              | ||||  |   | |||||  | |       | | || | |  | ||| | ||||
Db       1117 ACACATATAGGAATACATACTGGGGAAAAACCCTACAAATGTAATCAATGTGATAAAGCC 1176

Qy        364 TTTTATAAACA---TCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTA 420
              ||||  ||||   ||||      ||  |||  | | |  |  |   ||  |  | |
Db       1177 TTTTCACAACACTTTCAACTCCAAAATCATATAAGAACACATACTAGAGAGATGCCTTAC 1236

Qy        421 ACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTA 480
              |  ||||| |||||| | ||||  ||     |||| |  |||  |        ||||    |
Db       1237 AAATGTAATCAATGTGACAAAGCCTTTTTACGACATAGAAATCTTCAGATTCATAGTAGA 1296

Qy        481 AAACATCATGGTG 493
```

```
                ||||||  ||| |
Db       1297 AAACATACTGGAG 1309
```

RESULT 13
BE161630
LOCUS BE161630 604 bp mRNA linear EST 21-JUN-2000
DEFINITION MR3-HT0446-260300-201-f06 HT0446 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE161630
VERSION BE161630.1 GI:8624351
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 604)
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., *O'Hare,M.J., Soares,F.*, Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=MR3-HT0446-260300-201-f06&t3=2000-03-26&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 583.

FEATURES Location/Qualifiers
source 1. .604
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_lib="HT0446"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI; Site_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

ORIGIN

Query Match 7.2%; Score 89.2; DB 10; Length 604;
Best Local Similarity 53.7%; Pred. No. 8.1e-07;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

```
Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |   || || || |||||  |||||  ||||   | ||  |||    |||     | ||
Db        102 AGAGAAACCTTACAAATGTGAAGAATGTGACAAAGTTTTTAGTCGCTAATCACACCTTGA 161

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                ||||   ||||  || | |   ||   |||||| |||||||  ||      ||||||
Db        162 AAGACATAGGAGAATTCATACTGGAGAGAAACCGTACAAATGTAAGGT------TTGTGA 215

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              || ||  ||       | |||||  ||| |  ||||||||||    ||  | | | | ||
Db        216 CAAGGCTTTCAGACGTGATTCACACCTGGCACAACATATTGTAATTCACACTGGAGAGAA 275
```

```
Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | ||| | |||||| || || | | || || ||
Db        276 ACCTTACAAGTGTAATGAGTGTGGCAAGACCTTTGTTCAAAATTCATCTCTTGTAATGCA 335

Qy        309 TGAAATCACCCATAC---AAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATT 365
              | | ||| |||||   | || || | |          |||| | | | |
Db        336 TAAGGTCATTCATACTGGAGAGAAATGTTACAAGTGTAATGAATGTGGCAAGGTTTTTAA 395

Qy        366 TTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
              | | |||     |  | | | |||||||| | |  |  || ||| | |  | |||
Db        396 TCACAAATCAAACCTTGCATGTCATCATAGACTTCATACTGGAGAGAAACCTTACAAGTG 455

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              |||  |||||   || |||||  | | ||   |||||  | | ||| ||| | |  ||
Db        456 TAATGAATGTGGCAAGGTTTTTAATTGAAAATCAAACCTTGAACATCATCATAGAGTTCA 515

Qy        486 TCATGGTGGATCTCCT 501
              |   || | |   |||
Db        516 TATCGGAGAAAAACCT 531
```

RESULT 14
BC032590
LOCUS BC032590 3609 bp mRNA linear HTC 19-NOV-2003
DEFINITION Homo sapiens cDNA clone IMAGE:5502691, containing frame-shift errors.
ACCESSION BC032590
VERSION BC032590.1 GI:21619671
KEYWORDS HTC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 3609)
AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E., Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 3609)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc_mgc@nhgri.nih.gov
Akhter,N., Ayele,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P., Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legaspi,R., Maduro,Q.L., Masiello,C., Maskeri,B., Mastrian,S.D.,McCloskey,J.C., McDowell,J., Pearson,R., Stantripop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 69 Row: g Column: 20
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 13430873
This clone has the following problem: frame shifted.
FEATURES Location/Qualifiers
source 1. .3609
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5502691"
/tissue_type="Eye, retinoblastoma"
/clone_lib="NIH_MGC_67"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
ORIGIN

Query Match 7.2%; Score 89.2; DB 11; Length 3609;
Best Local Similarity 53.7%; Pred. No. 6.7e-07;
Matches 234; Conservative 0; Mismatches 193; Indels 9; Gaps 2;

```
Qy         69 AAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATCATTATTAGA 128
              | || |   || || || |||||  ||||| ||||  | || |||   ||||    | ||
Db        587 AGAGAAACCTTACAAATGTGAAGAATGTGACAAAGTTTTTAGTCGCTAATCACACCTTGA 646

Qy        129 GCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGACGATTGTGA 188
                ||||   ||||  || | |   ||  |||||| |||||||  ||       ||||||
Db        647 AAGACATAGGAGAATTCATACTGGAGAGAAACCGTACAAATGTAAGGT------TTGTGA 700

Qy        189 TAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTCCGAAAAAAA 248
              || || ||      | | ||||| ||| |  ||||||||||   || | | | | ||
Db        701 CAAGGCTTTCAGACGTGATTCACACCTGGCACAACATATTGTAATTCACACTGGAGAGAA 760

Qy        249 ACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTTGAAAAGACA 308
              ||| | | | |||  | |||||| ||  ||  |  |  | || ||
Db        761 ACCTTACAAGTGTAATGAGTGTGGCAAGACCTTTGTTCAAAATTCATCTCTTGTAATGCA 820

Qy        309 TGAAATCACCCATAC---AAAGTCATTTAAATGTACATTTGAAAATTGTCAAGAAGCATT 365
              | | |||  |||||   | ||  || | |        |||| | | | |
Db        821 TAAGGTCATTCATACTGGAGAGAAATGTTACAAGTGTAATGAATGTGGCAAGGTTTTTAA 880

Qy        366 TTATAAACATCAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAACATTAACGTG 425
              | | |||     |  | | | |||||||| | |  |  || ||| | |  | |||
Db        881 TCACAAATCAAACCTTGCATGTCATCATAGACTTCATACTGGAGAGAAACCTTACAAGTG 940

Qy        426 TAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAAATTAAAACA 485
              |||  |||||   || |||||  | | ||   |||||  | | ||| ||| | |  ||
Db        941 TAATGAATGTGGCAAGGTTTTTAATTGAAAATCAAACCTTGAACATCATCATAGAGTTCA 1000

Qy        486 TCATGGTGGATCTCCT 501
              |   || | |   |||
Db       1001 TATCGGAGAAAAACCT 1016
```

RESULT 15
AL705393
LOCUS AL705393 535 bp mRNA linear EST 04-SEP-2003
DEFINITION DKFZp686M1835_r1 686 (synonym: hlcc3) Homo sapiens cDNA clone DKFZp686M1835 5', mRNA sequence.
ACCESSION AL705393
VERSION AL705393.1 GI:19688748

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 535)
AUTHORS Ottenwaelder,B., Obermaier,B., Mewes,W., Mewes,H.W., Weil,B. and Wiemann,S.
TITLE EST (Ottenwaelder,B., Obermaier,B., Mewes,H.W., Weil,B. and Wiemann,S.)
JOURNAL Unpublished (2001)
COMMENT Contact: MIPS
MIPS
Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
sequenced by MediGenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. No s1 sequence available.
This clone (DKFZp686M1835) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES Location/Qualifiers
source 1. .535
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686M1835"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlcc3)"
/note="Vector: pTriplEx2; Site_1: SfiIA; Site_2: SfiIB; cDNA-collection"

ORIGIN

Query Match 7.2%; Score 88.6; DB 9; Length 535;
Best Local Similarity 53.3%; Pred. No. 1.1e-06;
Matches 237; Conservative 0; Mismatches 199; Indels 9; Gaps 2;

```
Qy         60 ACGTCCCAAAAAGTATATTTGCACATATGAAGGGTGTGATAAAGCCTATAATCGACCATC 119
              || | |   |||| |   || || || | |||  ||||| ||  | | |      | ||
Db         86 ACATACTGGAAAGAAATCTTTCAAATGTAAAGAATGTGAAAAGTCATTTTGCATGCTTTC 145

Qy        120 ATTATTAGAGCAACATTTAAGAACCCACAGTAATGATCGACCGTATAAATGTACAGTGGA 179
              |   ||||  ||||||  |||||  || |||    ||   ||| || |||||||
Db        146 ACACTTAGCTCAACATAAAAGAATTCATAGTGGAGAGAAACCCTACAAATGTAAA----- 200

Qy        180 CGATTGTGATAAAGCATTTTTCAGAAAATCACATTTGGAAACACATATTGTATCACATTC 239
               || ||||  ||||| | |      | ||| |   |    |||||||     |   ||| |
Db        201 -GAATGTGGGAAAGCCTATAATGAGACCTCAAACCTTTCTACACATAAAAGAATTCATAC 259

Qy        240 CGAAAAAAAACCATTCCATTGTTCAGTGTGTGGTAAAGGGGTTAATTCTCGACAACACTT 299
               | ||| ||||| | | | ||   || ||||||  ||||   ||||    |   ||||
Db        260 TGGAAAGAAACCCTACAAATGCGAAGAGTGTGGAAAAGCCTTTAACCGGCTCTCACACCT 319

Qy        300 GAAAAGACATGAAATCACCCATACAAAGTCATTTAAATGTACATTTGAAAATTGTCAAGA 359
               |  | |||| | || |  |||||              | | || |||  | |||   |
Db        320 TACTACACATAAGATAATTCATACTGGAAAGAAACCCTACAAATGTGAGGAGTGTGGCAA 379

Qy        360 AGCATTTTATAAACAT---CAATCTTTAAGACATCATATATTATCTGTTCATGAAAAAAC 416
              ||| ||| |  ||  |    |   |  | |||| | | | |   |  |   || ||| |
Db        380 AGCTTTTAACCAATCTGCAAACCTTACTACACATAAGAGAATTCATACTGGAGAGAAACC 439

Qy        417 ATTAACGTGTAAACAATGTAATAAAGTTTTCACTCGACCTTCAAAATTAGCACAACATAA 476
               |  |  ||| || |||||   | || ||| |  |   | ||||   |  |  ||||||
Db        440 CTACAAATGTGAAGAATGTGGCAGAGCTTTTAGCCAGTCCTCAACCCTTACTGCACATAA 499

Qy        477 ATTAAAACATCATGGTGGATCTCCT 501
               |||  |||  ||| |     |||
Db        500 GATAATTCATGCTGGAGAGAAACCT 524
```

Search completed: May 9, 2004, 10:57:43
Job time : 3623.38 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:14:51 ; Search time 58 Seconds
(without alignments)
2007.062 Million cell updates/sec

Title: US-09-831-804-3
Perfect score: 2229
Sequence: 1 MSESDETKSISSLISSSSSS.....PLVKKARMDLLPNETSVISR 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2229	100.0	412	3	AAY93316	Aay93316 A transcr
2	2225	99.8	412	5	ABP73656	Abp73656 Candida a
3	445	20.0	564	6	ABJ25655	Abj25655 Aspergill
4	445	20.0	564	6	ABJ26255	Abj26255 Aspergill
5	416.5	18.7	1212	4	ABG00399	Abg00399 Novel hum
6	403	18.1	1230	4	AAU30831	Aau30831 Novel hum
7	400	17.9	809	4	AAM38689	Aam38689 Human pol
8	398	17.9	563	6	ABU96690	Abu96690 Human nuc
9	395.5	17.7	365	3	AAY93317	Aay93317 A human t
10	395.5	17.7	409	4	AAG75181	Aag75181 Human col
11	395.5	17.7	423	2	AAR91305	Aar91305 Transcrip
12	395	17.7	622	4	AAM78947	Aam78947 Human pro
13	395	17.7	631	4	AAM79931	Aam79931 Human pro
14	395	17.7	632	4	ABG18386	Abg18386 Novel hum
15	395	17.7	719	4	ABG16953	Abg16953 Novel hum
16	393.5	17.7	1472	4	AAU31578	Aau31578 Novel hum
17	393	17.6	517	4	ABG03375	Abg03375 Novel hum
18	392.5	17.6	1050	4	ABG09685	Abg09685 Novel hum
19	392.5	17.6	1050	7	ADC32995	Adc32995 Human nov
20	392	17.6	555	6	ADA54763	Ada54763 Human pro
21	390.5	17.5	878	4	AAU29528	Aau29528 Novel hum
22	390.5	17.5	1520	4	ABG27130	Abg27130 Novel hum
23	390	17.5	540	7	ADB65594	Adb65594 Human pro
24	390	17.5	576	6	ADA54797	Ada54797 Human pro
25	390	17.5	620	6	ABU96725	Abu96725 Human nuc
26	388.5	17.4	501	4	ABB71369	Abb71369 Drosophil
27	388	17.4	568	5	ABB79872	Abb79872 TRAF6-inh
28	387.5	17.4	872	7	ADC37535	Adc37535 Human nuc
29	386	17.3	727	4	ABG16954	Abg16954 Novel hum
30	385.5	17.3	474	6	ABU11782	Abu11782 Human MDD
31	385	17.3	577	5	AAE14680	Aae14680 Human tra
32	385	17.3	770	4	ABG10091	Abg10091 Novel hum
33	385	17.3	770	4	ABG14920	Abg14920 Novel hum
34	385	17.3	803	4	ABG20106	Abg20106 Novel hum
35	380	17.0	615	6	AAE37047	Aae37047 Human nuc
36	380	17.0	959	4	ABG01956	Abg01956 Novel hum
37	379	17.0	817	4	AAM40475	Aam40475 Human pol
38	378	17.0	613	4	AAB95862	Aab95862 Human pro
39	378	17.0	755	4	AAM40916	Aam40916 Human pol
40	378	17.0	907	4	AAB94428	Aab94428 Human pro
41	377	16.9	518	6	ABR41469	Abr41469 Human DIT
42	376	16.9	553	7	ADC31307	Adc31307 Human nov
43	376	16.9	582	6	ABR41407	Abr41407 Human DIT
44	376	16.9	675	5	ABP55423	Abp55423 Human zin
45	376	16.9	678	7	ADC31847	Adc31847 Human nov

ALIGNMENTS

RESULT 1
AAY93316
ID AAY93316 standard; protein; 412 AA.
XX
AC AAY93316;
XX
DT 04-SEP-2000 (first entry)
XX
DE A transcription factor designated CATFIIIA.
XX
KW Transcription factor; CATFIIIA; DNA-binding protein;
KW ribosomal RNA 5S gene; fungal infection.
XX
OS Candida albicans.
XX
FH Key Location/Qualifiers
FT Misc-difference 193
FT /note= "Ser encoded by CTG"
FT Misc-difference 339
FT /note= "Ser encoded by CTG"
XX
PN WO200028037-A1.
XX
PD 18-MAY-2000.
XX
PF 09-NOV-1999; 99WO-FR002739.
XX
PR 10-NOV-1998; 98FR-00014147.
XX
PA (HMRI) HOECHST MARION ROUSSEL.
XX
PI Bordon-Pallier F, Camier S, Sentenac A;
XX
DR WPI; 2000-376549/32.
DR N-PSDB; AAA15398.
XX
PT New nucleic acid encoding Candida albicans transcription factor, useful
PT e.g. in screening for antimycotic agents and for immunization.
XX
PS Claim 12; Page 35-36; 45pp; French.
XX
CC The present sequence represents a Candida albicans transcription factor,
CC designated CATFIIIA. The polypeptide is a DNA-binding protein, which is
CC involved in initiating transcription of the ribosomal RNA 5S gene. The
CC polynucleotide is used to screen for its specific inhibitors, potentially
CC useful as antimycotic agents, to raise an antibody response that is
CC protective against fungal infection and to raise antibodies. Such
CC antibodies, as well as the polypeptides and polynucleotides are used in

CC compositions for diagnosing and treating fungal infections, e.g. by
CC detecting polymorphisms and mutations
XX
SQ Sequence 412 AA;

Query Match 100.0%; Score 2229; DB 3; Length 412;
Best Local Similarity 100.0%; Pred. No. 1e-167;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVD 60

Qy         61 DCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 120

Qy        121 AFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNF 180

Qy        181 QTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAK 240

Qy        241 KNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEE 300

Qy        301 DSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREY 360

Qy        361 DLRRHLKWHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DLRRHLKWHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
```

RESULT 2
ABP73656
ID ABP73656 standard; protein; 412 AA.
XX
AC ABP73656;
XX
DT 30-JAN-2003 (first entry)
XX
DE Candida albicans essential protein SEQ ID NO 7493.
XX
KW Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW signal transduction; DNA replication; cell division; growth;
KW proliferation; Candida albicans; fungicide; antifungal.
XX
OS Candida albicans.
XX
PN WO200253728-A2.
XX
PD 11-JUL-2002.
XX
PF 26-DEC-2001; 2001WO-US049486.
XX
PR 29-DEC-2000; 2000US-0259128P.
PR 20-FEB-2001; 2001US-00792024.
PR 22-AUG-2001; 2001US-0314050P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX
DR WPI; 2002-566694/60.
DR N-PSDB; ABZ32206.
XX
PT Constructing strains for identifying gene products as effective targets
PT for therapeutic intervention, by inactivating in the strain one allele of
PT a gene and placing other allele of the gene under conditional expression.
XX
PS Claim 44; SEQ ID NO 7493; 167pp + Sequence Listing; English.
XX
CC The invention relates to constructing (M1) a strain of diploid fungal
CC cells in which both alleles of a gene are modified, comprising modifying
CC one allele by insertion or replacement by a cassette having an
CC expressible selectable marker and modifying other allele by
CC recombination, of a promoter replacement fragment with a heterologous
CC promoter, so that expression of the second allele is regulated by the
CC promoter. (M1) is useful for constructing a strain of diploid fungal
CC cells in which both alleles of a gene are modified. The diploid fungal
CC cells having both alleles modified are useful for identifying a gene that
CC is essential to the survival or growth of a fungus, a gene that
CC contributes to the virulence and/or pathogenicity of a fungus, a gene
CC that contributes to the resistance of a diploid fungus to an antifungal
CC agent, an antifungal agent that inhibits the growth of a diploid fungus
CC and for identifying a therapeutic agent for treatment of a mammalian
CC disease. (M1) is useful for identifying a compound which modulates the
CC activity of a gene product, preferably enzymatic activity, carbon
CC compound catabolism, biosynthetic, transporter, transcriptional,
CC translational, signal transduction, DNA replication and cell division
CC activity. The method is useful for identifying a compound having the
CC ability to inhibit growth or proliferation of C. albicans cells and for
CC treating infection by C. albicans. The present sequence is that of an
CC essential Candida albicans protein used in the method of the invention.
CC Note: The sequence data for this patent is not represented in the printed
CC specification but is based on sequence information supplied to Derwent by
CC the European Patent Office
XX
SQ Sequence 412 AA;

Query Match 99.8%; Score 2225; DB 5; Length 412;
Best Local Similarity 99.8%; Pred. No. 2.1e-167;
Matches 411; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MSESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVD 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||:
Db          1 MSESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVE 60

Qy         61 DCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 DCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 120

Qy        121 AFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNF 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 AFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNF 180

Qy        181 QTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAK 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAK 240

Qy        241 KNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEE 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 KNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEE 300

Qy        301 DSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 DSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREY 360

Qy        361 DLRRHLKWHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 DLRRHLKWHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412
```

RESULT 3
ABJ25655
ID ABJ25655 standard; protein; 564 AA.
XX
AC ABJ25655;
XX
DT 16-APR-2003 (first entry)
XX

DE Aspergillus fumigatus essential gene protein #313.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013142.
XX
PR 23-APR-2001; 2001US-0285697P.
PR 27-APR-2001; 2001US-0287066P.
PR 05-JUN-2001; 2001US-0295890P.
PR 09-JUL-2001; 2001US-0303899P.
PR 31-AUG-2001; 2001US-0316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS Disclosure; Page; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and
CC making oligomers for attachment to a nucleic acid array for examination
CC of expression patterns, for raising anti-protein antibodies, as an
CC antigen to raise anti-DNA antibodies or to elicit another immune
CC response, and for identifying polynucleotides encoding the other protein
CC with which binding occurs or to identify inhibitors of the binding
CC interaction. The polypeptides may be used to raise antibodies or to
CC elicit immune response, as a reagent in assays designed to quantitatively
CC determine levels of the protein in biological fluids, as a marker for
CC host tissues in which pathogenic organism invade or reside, and to
CC isolate correlative receptors or ligands in the case or virulence
CC factors. This sequence represents a protein of one of the essential genes
CC of Aspergillus fumigatus of the invention
XX
SQ Sequence 564 AA;

Query Match 20.0%; Score 445; DB 6; Length 564;
Best Local Similarity 30.3%; Pred. No. 1.8e-26;
Matches 125; Conservative 62; Mismatches 164; Indels 62; Gaps 21;

```
Qy         14 ISSSSSSRP---KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKS 70
              ||::|:  |    | : | ::|| ||:|||: |::|||:|:|:| :||| ::||| | | |
Db         71 ISTTSAKYPSELKTHRCPFDGCTKAFNRPARLQEHLRSHNNERIFKCTFEECDKTFLRAS 130

Qy         71 HLETHIVS-HSEKKPFHCS--VCGKGVNSRQHLKRHEITH--TKSFKCT-FENCQEAFYK 124
              ||  || | |:  : : |    |||   :   |:||   |     ::|| :  | | | |
Db        131 HLNHHIKSAHTGVRDYVCDRPGCGKSFVTGSRLRRHLAAHDGRDKYRCTEYPPCNETFRK 190

Qy        125 HQSLRHHILSVH--EKTLTC--------KQCNKVFTRPSKLAQHKLKHHGGSPAYQCDH- 173
              | :|: ||:: |  :|   |         ::|   |     |  |: : |     : |
Db        191 HSTLQKHIMTAHLKQKPFQCPHTDPSTGQKCTMAFDTAGHLRAHESRIH-TEKRFSCTEC 249

Qy        174 ----PGCFKNFQTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHM-LSHDDSTM--IK 226
                  |   | |:::|| ||:  ||  :|| |  |  :  |  |: ::| | ::   |
Db        250 SQHAEGAEATFPTYALLQAHIRSVHPP-QCPNCALTCATSRELRRHLEVAHGDVSLEERK 308

Qy        227 IWTCDY--CDVGKFAKKNELVEHYNIFHDGNI-----PDDLLKETEVKKLENLLDQ---- 275
              |: |   ||  | || |  |  | |          ||    :|   |  |
Db        309 IFPCTVPGCD-RSFTKKGNLTVHIRTVHQGEKRFVCGETDLSSSKKVSGWNN--DNGCGK 365

Qy        276 --GSKLNNLHELETEKLKVEEDEEDEEDSL----DEKRSDVRSDSMSAQRSIKSFTASLE 329
                ||||     : |  |  :  : :    |    | : |   |  :||        |:|
Db        366 RYGSKLALEEHIRTAHLGYQNAKAERRQRLGITRDRQHSTATSPGVSA-------LAALT 418

Qy        330 GSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHL--KWHDDNLQRIESFL 380
              |     :   :|: | |   :|   | |:|||  |:  | |    :  : ||
Db        419 GEGYAEE----TGRHIACLVESCPHRFHRDYDLWVHMSGKHHFSEEETRDLFL 467
```

RESULT 4
ABJ26255
ID ABJ26255 standard; protein; 564 AA.
XX
AC ABJ26255;
XX
DT 16-APR-2003 (first entry)
XX
DE Aspergillus fumigatus essential gene protein #913.
XX
KW Fungicide; cytostatic; essential gene; Aspergillus fumigatus; infection;
KW cancer; contamination; biofilm; antibody; immune response.
XX
OS Aspergillus fumigatus.
XX
PN WO200286090-A2.
XX
PD 31-OCT-2002.
XX
PF 23-APR-2002; 2002WO-US013142.
XX
PR 23-APR-2001; 2001US-0285697P.
PR 27-APR-2001; 2001US-0287066P.
PR 05-JUN-2001; 2001US-0295890P.
PR 09-JUL-2001; 2001US-0303899P.
PR 31-AUG-2001; 2001US-0316362P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Jiang B, Tishkoff D, Zamudio C, Eroshkin AM, Hu W, Lemieux SM;
XX
DR WPI; 2003-093124/08.
XX
PT New purified or isolated nucleic acids of essential genes of Aspergillus
PT fumigatus, useful for treating or preventing infections by A. fumigatus,
PT or for treating a non-infectious disease in a subject e.g. cancer.
XX
PS Disclosure; Page; 175pp; English.
XX
CC The invention relates to novel purified or isolated nucleic acids of
CC essential genes of Aspergillus fumigatus. The isolated nucleic acids of
CC the invention are used to treat or prevent infections by a pathogenic
CC organism such as A. fumigatus, to treat a non-infectious disease in a
CC subject (e.g. cancer), to prevent or contain contamination of an object
CC by A. fumigatus, or to prevent or inhibit formation on a surface of a
CC biofilm comprising A. fumigatus. The polynucleotides are useful for
CC expressing recombinant protein for characterisation, screening or
CC therapeutic use, as markers for host tissues in which the pathogenic
CC organisms invade or reside, for comparing with the DNA sequence of A.
CC fumigatus to identify duplicated genes or paralogues having the same or
CC similar biochemical activity and/or function, for comparing with DNA
CC sequences of other related or distant pathogenic organisms to identify
CC potential orthologous essential or virulence genes, for selecting and

```
CC   making oligomers for attachment to a nucleic acid array for examination
CC   of expression patterns, for raising anti-protein antibodies, as an
CC   antigen to raise anti-DNA antibodies or to elicit another immune
CC   response, and for identifying polynucleotides encoding the other protein
CC   with which binding occurs or to identify inhibitors of the binding
CC   interaction. The polypeptides may be used to raise antibodies or to
CC   elicit immune response, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, as a marker for
CC   host tissues in which pathogenic organism invade or reside, and to
CC   isolate correlative receptors or ligands in the case or virulence
CC   factors. This sequence represents a protein of one of the essential genes
CC   of Aspergillus fumigatus of the invention
XX
SQ   Sequence 564 AA;

  Query Match             20.0%;  Score 445;  DB 6;  Length 564;
  Best Local Similarity   30.3%;  Pred. No. 1.8e-26;
  Matches  125;  Conservative   62;  Mismatches  164;  Indels   62;  Gaps   21;

Qy         14 ISSSSSSRP---KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKS 70
Db         71 ISTTSAKYPSELKTHRCPFDGCTKAFNRPARLQEHLRSHNNERIFKCTFEECDKTFLRAS 130

Qy         71 HLETHIVS-HSEKKPFHCS--VCGKGVNSRQHLKRHEITH--TKSFKCT-FENCQEAFYK 124
Db        131 HLNHHIKSAHTGVRDYVCDRPGCGKSFVTGSRLRRHLAAHDGRDKYRCTEYPPCNETFRK 190

Qy        125 HQSLRHHILSVH--EKTLTC--------KQCNKVFTRPSKLAQHKLKHHGGSPAYQCDH- 173
Db        191 HSTLQKHIMTAHLKQKPFQCPHTDPSTGQKCTMAFDTAGHLRAHESRIH-TEKRFSCTEC 249

Qy        174 ----PGCFKNFQTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHM-LSHDDSTM--IK 226
Db        250 SQHAEGAEATFPTYALLQAHIRSVHPP-QCPNCALTCATSRELRRHLEVAHGDVSLEERK 308

Qy        227 IWTCDY--CDVGKFAKKNELVEHYNIFHDGNI-----PDDLLKETEVKKLENLLDQ---- 275
Db        309 IFPCTVPGCD-RSFTKKGNLTVHIRTVHQGEKRFVCGETDLSSSKKVSGWNN--DNGCGK 365

Qy        276 --GSKLNNLHELETEKLKVEEDEEDEEDSL----DEKRSDVRSDSMSAQRSIKSFTASLE 329
Db        366 RYGSKLALEEHIRTAHLGYQNAKAERRQRLGITRDRQHSTATSPGVSA-------LAALT 418

Qy        330 GSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHL--KWHDDNLQRIESPL 380
Db        419 GEGYAEE----TGRHIACLVESCPHRFHRDYDLWVHMSGKHHFSEEETRDLFL 467


RESULT 5
ABG00399
ID   ABG00399 standard; protein; 1212 AA.
XX
AC   ABG00399;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #390.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS64586.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 30758; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1212 AA;

  Query Match             18.7%;  Score 416.5;  DB 4;  Length 1212;
  Best Local Similarity   32.6%;  Pred. No. 9.1e-24;
  Matches  116;  Conservative   45;  Mismatches  148;  Indels   47;  Gaps   15;

Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
Db        861 TREKTYKC--EECGKAFSQPSHLTTHKRMHTGEKPYKC--EECGKAFSQSSTLTTHKIIH 916

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
Db        917 TGEKPYKCEECGKAFRKSSTLTEHKIIHTGEKPYKC--EECGKAFSQSSTLTRHTRMHTG 974

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK 196
Db        975 EKPYKCEECGKAFNRSSKLTTHKIIHTGEKP-YKCEE--CGKAFISSSTLNGH-KRIHTR 1030

Qy        197 ---LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHD 253
Db       1031 EKPYKCEECGKAFSQSSTLTRHKRLH---TGEKPYKCGEC--GKAFKESSALTKHKIIHT 1085

Qy        254 GNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSD 313
Db       1086 GE------KPYKCEKCGKAFNQSSILTNHKKIHTITPKIH---------------TREK 1123

Qy        314 SMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
Db       1124 PYKYKECGKSFNRSSTFTK--HKVIHTGVKLYKC--EECGKSFFWSSALTRHKKIH 1175


RESULT 6
AAU30831
ID   AAU30831 standard; protein; 1230 AA.
XX
AC   AAU30831;
XX
DT   18-DEC-2001  (first entry)
```

XX
DE Novel human secreted protein #1322.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
PS Claim 20; Page 355; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
SQ Sequence 1230 AA;

Query Match 18.1%; Score 403; DB 4; Length 1230;
Best Local Similarity 32.9%; Pred. No. 1.1e-22;
Matches 118; Conservative 45; Mismatches 144; Indels 52; Gaps 17;

```
Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
              :| | | |  | | ||:::|| | | | |: ::|||| ::| ||| : | | || : |
Db        891 TREKPYKC--EECGKAFSQPSHLTTHKRMHTGEKPYKC--EECGKAFSQSSTLTTHKIIH 946

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
              : :||: |  |||       |  |:| ||  | :||  | | :|| :  :| ||  :
Db        947 TGEKPYKCEECGKAFRKSSTLTEHKIIHTGEKPYKC--EECGKAFSQSSTLTRHTRMHTG 1004

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK 196
              ||   |::| | | | ||| ||: | |  | |:|:   | | | : | |  | |: | :
Db       1005 EKPYKCEECGKAFNRSSKLTTHKIIHTGEKP-YKCEE--CGKAFISSSTLNGH-KRIHTR 1060

Qy        197 LKCPKCGKGCVGK------KGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNI 250
              |  || :|| ||        |: |   |   | | : | |  ||  |::  :  : |
Db       1061 EKPYKC-EGC-GKAFSQSFNTLTGHKRLH---TGEKPYKCGEC--GKAFKESSALTKHKI 1113

Qy        251 FHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDV 310
              | |      | : :|     :| | | |  :: |   |:
Db       1114 IHTGE------KPYKCEKCCKAFNQSSILTNHKKIHTITPKIH----------------T 1151
```

```
Qy        311 RSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
              |      :  ||| |  :|  |:|    |  |   | : |     | || | |
Db       1152 REKPYKYKECGKSFNRSSTFTK--HKVIHTGVKLYKC--EECGKSFFWSSALTRHKKIH 1206
```

RESULT 7
AAM38689
ID AAM38689 standard; protein; 809 AA.
XX
AC AAM38689;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 1834.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US034263.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
PR 20-JUN-2000; 2000US-00598042.
PR 19-JUL-2000; 2000US-00620312.
PR 03-AUG-2000; 2000US-00653450.
PR 14-SEP-2000; 2000US-00662191.
PR 19-OCT-2000; 2000US-00693036.
PR 29-NOV-2000; 2000US-00727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J, Zhao QA;
PI Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR N-PSDB; AAI57845.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders such
PT as central nervous system injuries.
XX
PS Example 3; SEQ ID NO 1834; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the
CC encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders. Note: The sequence data for this patent did not form
CC part of the printed specification
XX
SQ Sequence 809 AA;

Query Match 17.9%; Score 400; DB 4; Length 809;

```
  Best Local Similarity   31.9%;  Pred. No. 1.1e-22;
  Matches  121;  Conservative   53;  Mismatches  171;  Indels   34;  Gaps   19;

Qy          2 SESDETKSISSLISSSSS--SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTV 59
              :|  |   | || ::      :  | | |  | | ||:   | | :|  ||: ::|||| 
Db        373 TECGEAFSRSSNLTKHKKIHTEKKPYKC--EECGKAFKWSSKLTEHKLTHTGEKPYKC-- 428

Qy         60 DDCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFEN 117
              ::| |||   | |  |   |: :||: | ||||  |   :|  |:  ||  | :||  | 
Db        429 EECGKAFNWPSTLTKHNRIHTGEKPYKCEVCGKAFNQFSNLTTHKRIHTAEKPYKC--EE 486

Qy        118 CQEAFYKHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGC 176
              | :|| :  :| :|  : : :|   |::| | |   ||| :||: | |  | |:|:  | 
Db        487 CGKAFSRSSNLTKHKKIHIEKKPYKCEECGKAFKWSSKLTEHKITHTGEKP-YKCEE--C 543

Qy        177 FKNFQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCD 234
              | |  :|:|  |  |        || :|||       |::|  |   |  | : |: | 
Db        544 GKAFNHFSILTKHKRIHTGEKPYKCEECGKAFTQSSNLTTHKKIH---TGEKFYKCEEC- 599

Qy        235 VGK-FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVE 293
              || | : :  |  | | |  |    |  |  |  |      :|  :|  | :  | |  | 
Db        600 -GKAFTQSSNLTTHKKI-HTGGKP---YKCEECGKAFNQFSTLTKHKIIH-TEEKPYKCE 653

Qy        294 E-DEEDEEDSLDEKRSDVRSDS--MSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKN 350
              |  :  :  |   |   : :       :  |:|   |  :  |  |:|    |   | | 
Db        654 ECGKAFKWSSTLTKHKIIHTGEKPYKCEECGKAF--KLSSTLSTHKIIHTGEKPYKCEK- 710

Qy        351 NCDRMFSREYDLRRHLKWH 369
              | : |:|  :|  | | | 
Db        711 -CGKAFNRPSNLIEHKKIH 728


RESULT 8
ABU96690
ID   ABU96690 standard; protein; 563 AA.
XX
AC   ABU96690;
XX
DT   25-JUL-2003  (first entry)
XX
DE   Human nucleic acid-associated protein (NAAP) #19.
XX
KW   Human; nucleic acid-associated protein; cytostatic; antiarteriosclerotic;
KW   anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;
KW   antiallergic; antiinflammatory; thyromimetic; gene therapy;
KW   cell proliferative disorder; cancer; atherosclerosis;
KW   neurological disorder; epilepsy; Huntington's disease; stroke;
KW   immune disorder; inflammatory disorder; AIDS; allergy;
KW   developmental disorder; Hypothyroidism; Cushing's syndrome; infection;
KW   protein-protein interaction; drug-target interaction;
KW   gene expression profile.
XX
OS   Homo sapiens.
XX
PN   WO2003023003-A2.
XX
PD   20-MAR-2003.
XX
PF   05-SEP-2002; 2002WO-US028540.
XX
PR   07-SEP-2001; 2001US-0317792P.
PR   07-SEP-2001; 2001US-0317912P.
PR   14-SEP-2001; 2001US-0322270P.
PR   21-SEP-2001; 2001US-0324040P.
PR   28-SEP-2001; 2001US-0326732P.
PR   19-OCT-2001; 2001US-0346716P.
PR   25-JAN-2002; 2002US-0351749P.
PR   22-FEB-2002; 2002US-0359498P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Tang YT,  Jackson JL,  Griffin JA,  Elliott VS,  Forsythe IJ;
PI   Becha SD,  Richardson TW,  Lee EA,  Sprague WW,  Emerling BM;
PI   Thangavelu K,  Warren BA,  Tran UK,  Yue H,  Xu Y,  Yue H,  Li JX;
PI   Hafalia AJA,  Sanjanwala B,  Marquis JP,  Gorvad AE,  Lee SY,  Ison CH;
PI   Baughn MR,  Chawla NK,  Nguyen DB,  Swarnakar A,  Zebarjadian Y,  Shah P;
PI   Thornton M,  Yao MG,  Khan FA,  Gandhi AR,  Yang J,  Kable AE;
PI   Burford N,  Ramkumar J;
XX
DR   WPI; 2003-313243/30.
DR   N-PSDB; ACA98938.
XX
PT   New human nucleic acid associated proteins (NAAP), useful for diagnosing,
PT   treating and preventing diseases or conditions associated with the
PT   aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or
PT   infections.
XX
PS   Claim 1; Page 255-257; 345pp; English.
XX
CC   The invention describes a novel human isolated nucleic acid-associated
CC   polypeptide (NAAP). The polypeptides and polynucleotides are useful in
CC   diagnosing, treating and preventing diseases or conditions associated
CC   with the decreased expression or overexpression of NAAP, such as cell
CC   proliferative (e.g. cancer, atherosclerosis), neurological (e.g.
CC   epilepsy, Huntington's disease, stroke), immune/inflammatory (e.g. AIDS,
CC   allergies) and developmental (e.g. Hypothyroidism, Cushing's syndrome)
CC   disorders, or infections. These are also useful in assessing the effects
CC   of exogenous compounds on the expression of nucleic acid and amino acid
CC   sequences of NAAP. The NAAP or its fragments are useful in screening
CC   compounds for effectiveness as agonist or antagonist of the polypeptides,
CC   or in altering the expression of the target polynucleotide and compounds
CC   that specifically bind to or modulate the activity of the polypeptide.
CC   The microarray is useful in monitoring or measuring protein-protein
CC   interactions, drug-target interactions, and gene expression profiles.
CC   This is the amino acid sequence of a novel human nucleic acid-associated
CC   protein (NAAP)
XX
SQ   Sequence 563 AA;

  Query Match             17.9%;  Score 398;  DB 6;  Length 563;
  Best Local Similarity   30.2%;  Pred. No. 9.3e-23;
  Matches  121;  Conservative   48;  Mismatches  167;  Indels   64;  Gaps   18;

Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
              |  | | |  : | ||||  | |  | | |:  :||||  ::| ||| | ||| || : | 
Db        196 SGEKPYKC--KECGKAYNETSNLSTHKRIHTGKKPYKC--EECGKAFNRLSHLTTHKIIH 251

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
              : |||: |  |||  |   :|  |:  ||  | :||  | |  || :  :|  | |:  
Db        252 TGKKPYKCEECGKAFNQSANLTTHKRIHTGEKPYKC--EECGRAFSQSSTLTAHKIIHAG 309

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCPKNFQTWSVLQFH--IKQSH 194
              ||   |::| | |:: | |  ||:  | |   |:|:   | | |  | |  |  |    
Db        310 EKPYKCEECGKAFSQSSTLTTHKI-IHTGEKFYKCEE--CGKAFSQLSHLTTHKRIHSGE 366

Qy        195 PKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDG 254
                 || :|||       |::|   |       | : |: |    |:: : |  |  | | | 
Db        367 KPYKCEECGKAFKQSSTLTTHKRIHAGE---KFYKCEVCSKA-FSRFSHLTTHKRI-HTG 421

Qy        255 NIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDS 314
                |    |  |  |  ||  | :   :|  | :  | ||            ::  :| : 
Db        422 EKP---YKCEECGKAFNLSSQLTTHKIIHTGE-KPYKCEEC----------GKAFNQSST 467

Qy        315 MSAQRSI-------------KSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYD 361
              :|  : |             |:|  |     :  |:|    |   |    | : |:  
Db        468 LSKHKVIHTGEKPYKYEECGKAFNQS--SHLTTHKMIHTGEKPYKC--EECGKAFNNSSI 523

Qy        362 LRRHLKWHD-------------DNLQRIESFLNSIEKEE 387
              | ||  |               ||: :|  :  :   |:  
Db        524 LNRHKMIHTGEKLYKPESCNNACDNIAKISKYKRNCAGEK 563


RESULT 9
AAY93317
ID   AAY93317 standard; protein; 365 AA.
XX
```

AC AAY93317;
XX
DT 04-SEP-2000 (first entry)
XX
DE A human transcription factor designated htfIIIA.
XX
KW Human; transcription factor; htfIIIA; DNA-binding protein; transcription;
KW ribosomal RNA 5S gene; transcriptional control; cancer.
XX
OS Homo sapiens.
XX
PN WO200028024-A1.
XX
PD 18-MAY-2000.
XX
PF 09-NOV-1999; 99WO-FR002738.
XX
PR 10-NOV-1998; 98FR-00014146.
XX
PA (HMRI) HOECHST MARION ROUSSEL.
XX
PI Bordon-Pallier F, Rocher C;
XX
DR WPI; 2000-387419/33.
DR N-PSDB; AAA15405.
XX
PT New nucleic acid encoding human transcription factor IIIA, useful for
PT treatment and diagnosis of cancer and inherited disease.
XX
PS Claim 10; Page 40-41; 49pp; French.
XX
CC The present sequence represents a human transcription factor designated
CC htfIIIA. The polypeptide is probabaly a DNA-binding protein probably
CC involved in initiating transcription of the gene for ribosomal RNA 5S and
CC maintaining the stability of transcription of other control genes. The
CC htfIIIA polynucleotides and polypeptides are used to make therapeutic or
CC diagnostic compositions for diseases associated with disorders of
CC transcriptional control, particularly cancer or other inherited diseases.
CC The htfIIIA polynucleotide can also be used to detect anomalies in gene
CC transcription, particularly for diagnosis of inherited disease, also for
CC studying diseases involving hTFIIIA
XX
SQ Sequence 365 AA;

Query Match 17.7%; Score 395.5; DB 3; Length 365;
Best Local Similarity 31.1%; Pred. No. 8.2e-23;
Matches 99; Conservative 54; Mismatches 96; Indels 69; Gaps 16;

```
Qy          8 KSISSL------ISSSSSSR--------PKKYICTYEGCDKAYNRPSLLEQHLRTHSNDR 53
              :|:|||       |::  ||          |:::||::  |    |::   |: ||  |: :|
Db          9 ESVSSLTIADAFIAAGESSAPTPPRPALPRRFICSFPDCSANYSKAWKLDAHLCKHTGER 68

Qy         54 PYKCTVDDCDKAFFRKSHLETHIVSHSEKKPFHCSV--CGKGVNSRQHLKRH-EITH--- 107
              |: |   :  | ||| |   ||   ||::|: :||| |:    | :  |:: :||:| |  |
Db         69 PFVCDYEGCGKAFIRDYHLSRHILTHTGEKPFVCAATGCDQKFNTKSNLKKHFERKHENQ 128

Qy        108 TKSFKCTFENCQEAFYKHQSLR-HHILSVHEKTLTCKQ--CNKVFTRPSKLAQHKLKHHG 164
               | : |:||:|:: | ||| |: |     :|    | |  | | |  |||| :|   | |
Db        129 QKQYICSFEDCKKTFKKHQQLKIHQCQHTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEG 188

Qy        165 GSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK---------------------------- 196
                 | |   ||    :||: |  |::::| :
Db        189 ----YVC-QKGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPER 243

Qy        197 --LKCPK--CGKGCVGKKGLSSHMLS-HDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNI 250
                :||:  ||:        | ||:|| |::|   : : |::   || || |  |  |  :
Db        244 DVCRCPREGCGRTYTTVFNLQSHILSFHEES---RPFVCEHAGCGKTFAMKQSLTRH-AV 299

Qy        251 FHDGNIPDDLLKETEVKK 268
               ||   ||    : :|||
Db        300 VHD---PDKKKMKLKVKK 314
```

RESULT 10
AAG75181
ID AAG75181 standard; protein; 409 AA.
XX
AC AAG75181;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:5945.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; chromosome 13.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US026524.
XX
PR 29-SEP-1999; 99US-0157137P.
PR 03-NOV-1999; 99US-0163280P.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR N-PSDB; AAH34586.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers.
XX
PS Claim 11; Page 7452-7454; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where the
CC proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene therapy
CC and vaccine production. N and P may be used in the prevention, diagnosis
CC and treatment of diseases associated with inappropriate P expression. For
CC example, N and P may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of P by expressing inactive proteins or to
CC supplement the patients own production of P. Additionally, N may be used
CC to produce the colon cancer-associated Ps, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the proteins. N and P
CC can be used in the prevention, diagnosis and treatment of colorectal
CC carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent
CC sequences used in the exemplification of the present invention. N.B.
CC Pages 666 to 682 and page 7053 of the sequence listing were missing at
CC time of publication, meaning no sequences are present for SEQ ID NO:1027
CC to 1052, 7921 and 7922
XX
SQ Sequence 409 AA;

Query Match 17.7%; Score 395.5; DB 4; Length 409;
Best Local Similarity 31.1%; Pred. No. 9.6e-23;
Matches 99; Conservative 54; Mismatches 96; Indels 69; Gaps 16;

```
Qy          8 KSISSL------ISSSSSSR--------PKKYICTYEGCDKAYNRPSLLEQHLRTHSNDR 53
              :|:|||       |::  ||          |:::||::  |    |::   |: ||  |: :|
Db         53 ESVSSLTIADAFIAAGESSAPTPPRPALPRRFICSFPDCSANYSKAWKLDAHLCKHTGER 112

Qy         54 PYKCTVDDCDKAFFRKSHLETHIVSHSEKKPFHCSV--CGKGVNSRQHLKRH-EITH--- 107
              |: |   :  | ||| |   ||   ||::|: :||| |:    | :  |:: :||:| |  |
Db        113 PFVCDYEGCGKAFIRDYHLSRHILTHTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQ 172

Qy        108 TKSFKCTFENCQEAFYKHQSLR-HHILSVHEKTLTCKQ--CNKVFTRPSKLAQHKLKHHG 164
               | : |:||:|:: | ||| |: |     :|    | |  | | |  |||| :|   | |
Db        173 QKQYICSFEDCKKTFKKHQQLKIHQCQHTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEG 232
```

```
Qy        165 GSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK---------------------------- 196
              | |   ||    :||: |  |::::| :
Db        233 ----YVC-QKGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPER 287

Qy        197 --LKCPK--CGKGCVGKKGLSSHMLS-HDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNI 250
                :||:  ||:       | ||:|| |::|   : : |::   || || |  |  :
Db        288 DVCRCPREGCGRTYTTVFNLQSHILSFHEES---RPFVCEHAGCGKTFAMKQSLTRH-AV 343

Qy        251 FHDGNIPDDLLKETEVKK 268
               ||   ||    : :|||
Db        344 VHD---PDKKKMKLKVKK 358


RESULT 11
AAR91305
ID   AAR91305 standard; protein; 423 AA.
XX
AC   AAR91305;
XX
DT   07-JUL-1996  (first entry)
XX
DE   Transcription factor-IIIA.
XX
KW   Human; transcription factor-IIIA; hTFIIIA; DNA binding protein; ribosome;
KW   zinc finger; diagnostic; probe; transcription control; antitumour;
KW   cancer; therapy.
XX
OS   Homo sapiens.
XX
PN   EP704526-A1.
XX
PD   03-APR-1996.
XX
PF   05-SEP-1995;   95EP-00113908.
XX
PR   05-SEP-1994;   94JP-00211022.
XX
PA   (SAKA ) OTSUKA PHARM CO LTD.
XX
PI   Fujiwara T,  Takeda S,  Shimada Y,  Ozaki K,  Shin S;
XX
DR   WPI; 1996-173033/18.
DR   N-PSDB; AAT14037, AAT14038.
XX
PT   Human Transcription Factor III A gene - useful in regulation of
PT   transcription and for diagnosis and treatment of e.g. cancer related
PT   diseases.
XX
PS   Claim 1; Page 8-10; 17pp; English.
XX
CC   The sequence represents human transcription factor-IIIA (hTFIIIA), a DNA
CC   binding protein which is necessary for the initiation of 5S RNA gene
CC   transcription, binding to an internal control region of the 5S gene. The
CC   protein contains 9 zinc finger domains, which are homologous to the C2H2
CC   finger domains of Xenopus TFIIIA, except for the 6th finger domain, which
CC   has only 3 amino acid residues between 2 cysteine residues, instead of 5
CC   amino acid residues for Xenopus TFIIIA. The protein (optionally in
CC   recombinant form) and encoding gene may be used in diagnosis,
CC   identification or therapy of hereditary diseases such as cancer, or other
CC   diseases resulting from abnormal transcriptional control, and to analyse
CC   the mechanisms involved in their activity
XX
SQ   Sequence 423 AA;

  Query Match           17.7%;  Score 395.5;  DB 2;  Length 423;
  Best Local Similarity   30.8%;  Pred. No. 1e-22;
  Matches   98;  Conservative   56;  Mismatches   95;  Indels   69;  Gaps   16;

Qy          8 KSISSL------ISSSSSSR--------PKKYICTYEGCDKAYNRPSLLEQHLRTHSNDR 53
              :|:|||      |::  ||         |:::||::  |   |::   |: ||  |: :|
Db         67 ESVSSLTIADAFIAAGESSAPTPPRPALPRRFICSFPDCSANYSKAWKLDAHLCKHTGER 126

Qy         54 PYKCTVDDCDKAFFRKSHLETHIVSHSEKKPFHCSV--CGKGVNSRQHLKRH-EITH--- 107
```

```
                |: |  : | ||| |  ||  ||::|: :||| |:   | :  |:: :||:| |  |
Db        127 PFVCDYEGCGKAFIRDYHLSRHILTHTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQ 186

Qy        108 TKSFKCTFENCQEAFYKHQSLR-HHILSVHEKTLTCKQ--CNKVFTRPSKLAQHKLKHHG 164
              | : |:||:|:: | ||| :: |   : :|    | |  | | |  |||| :|   | |
Db        187 QKQYICSFEDCKKTFKKHQQMKIHQCQNTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEG 246

Qy        165 GSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK---------------------------- 196
                | |   ||    :||: |  |::::| :
Db        247 ----YVC-QKGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPER 301

Qy        197 --LKCPK--CGKGCVGKKGLSSHMLS-HDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNI 250
                :||:  ||:       | ||:|| |::|   : : |::   || || |  |  :
Db        302 DVCRCPREGCGRTYTTVFNLQSHILSFHEES---RPFVCEHAGCGKTFAMKQSLTRH-AV 357

Qy        251 FHDGNIPDDLLKETEVKK 268
               ||   ||    : :|||
Db        358 VHD---PDKKKMKLKVKK 372


RESULT 12
AAM78947
ID   AAM78947 standard; protein; 622 AA.
XX
AC   AAM78947;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 1609.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00598075.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y;
PI   Ma Y,  Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK52080.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
PS   Claim 20; Page 3945-3946; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
```

```
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 622 AA;

  Query Match             17.7%;  Score 395;  DB 4;  Length 622;
  Best Local Similarity   29.7%;  Pred. No. 1.8e-22;
  Matches  123;  Conservative   42;  Mismatches  157;  Indels   92;  Gaps   20;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
Db        201 KPYIC--EECGKAFKYSSALNTHKRIHTGEKPYKC--DKCDKAFIASSTLSKHEIIHTGK 256

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHH--------- 131
Db        257 KPYKCEECGKAFNQSSTLTKHKKIHTGEKPYKC--EECGKAFNQSSTLTKHKKIHTGEKP 314

Qy        132 --------------ILSVH------EKTLTCKQCNKVFTRPSKLAQHKLKH--------- 162
Db        315 YVCEECGKAFKYSRILTTHKRIHTGEKPYKCNKCGKAFIASSTLSRHEFIHMGKKHYKCE 374

Qy        163 ------------------HGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSH---PKLKCPK 201
Db        375 ECGKAFIWSSVLTRHKRVHTGEKPYKCEE--CGKAFKYSSTLSSH-KRSHTGEKPYKCEE 431

Qy        202 CGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGNIPDDL 260
Db        432 CGKAFVASSTLSKHEIIH---TGKKPYKCEEC--GKAFNQSSSLTKHKKI-HTGEKP--- 482

Qy        261 LKETEVKKLENLLDQGSKLNNLHELET--EKLKVEEDEEDEEDS---LDEKRSDVRSDSM 315
Db        483 YKCEECGKAFN---QSSSLTKHKKIHTGEKPYKCEECGKAFNQSSTLIKHKKIHTREKPY 539

Qy        316 SAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
Db        540 KCEECGKAF--HLSTHLTTHKILHTGEKPYRC--RECGKAFNHSATLSSHKKIH 589


RESULT 13
AAM79931
ID   AAM79931 standard; protein; 631 AA.
XX
AC   AAM79931;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human protein SEQ ID NO 3577.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00598075.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT,  Asundi V,  Zhou P,  Xu C,  Cao Y;
PI   Ma Y,  Zhao QA,  Wang D,  Wang J,  Zhang J,  Ren F,  Chen R,  Wang ZW;
PI   Xue AJ,  Yang Y,  Wejhrman T,  Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK53064.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
PS   Claim 20; Page 388; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 631 AA;

  Query Match             17.7%;  Score 395;  DB 4;  Length 631;
  Best Local Similarity   29.7%;  Pred. No. 1.9e-22;
  Matches  123;  Conservative   42;  Mismatches  157;  Indels   92;  Gaps   20;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
Db        210 KPYIC--EECGKAFKYSSALNTHKRIHTGEKPYKC--DKCDKAFIASSTLSKHEIIHTGK 265

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHH--------- 131
Db        266 KPYKCEECGKAFNQSSTLTKHKKIHTGEKPYKC--EECGKAFNQSSTLTKHKKIHTGEKP 323

Qy        132 --------------ILSVH------EKTLTCKQCNKVFTRPSKLAQHKLKH--------- 162
Db        324 YVCEECGKAFKYSRILTTHKRIHTGEKPYKCNKCGKAFIASSTLSRHEFIHMGKKHYKCE 383

Qy        163 ------------------HGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSH---PKLKCPK 201
Db        384 ECGKAFIWSSVLTRHKRVHTGEKPYKCEE--CGKAFKYSSTLSSH-KRSHTGEKPYKCEE 440

Qy        202 CGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGNIPDDL 260
Db        441 CGKAFVASSTLSKHEIIH---TGKKPYKCEEC--GKAFNQSSSLTKHKKI-HTGEKP--- 491

Qy        261 LKETEVKKLENLLDQGSKLNNLHELET--EKLKVEEDEEDEEDS---LDEKRSDVRSDSM 315
Db        492 YKCEECGKAFN---QSSSLTKHKKIHTGEKPYKCEECGKAFNQSSTLIKHKKIHTREKPY 548

Qy        316 SAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
Db        549 KCEECGKAF--HLSTHLTTHKILHTGEKPYRC--RECGKAFNHSATLSSHKKIH 598


RESULT 14
ABG18386
ID   ABG18386 standard; protein; 632 AA.
XX
AC   ABG18386;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #18377.
```

```
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS82573.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 48745; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 632 AA;

  Query Match             17.7%;  Score 395;  DB 4;  Length 632;
  Best Local Similarity   29.7%;  Pred. No. 1.9e-22;
  Matches  123;  Conservative   42;  Mismatches  157;  Indels   92;  Gaps   20;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
              |  |||  | | ||:   | |  | | |: ::||||  | |||||   | |  | : |: |
Db        211 KPYIC--EECGKAFKYSSALNTHKRIHTGEKPYKC--DKCDKAFIASSTLSKHEIIHTGK 266

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHH--------- 131
              ||: |  |||  |    | :|:  ||  | :||  | | :|| :  :|  |
Db        267 KPYKCEECGKAFNQSSTLTKHKKIHTGEKPYKC--EECGKAFNQSSTLTKHKKIHTGEKP 324

Qy        132 --------------ILSVH------EKTLTCKQCNKVFTRPSKLAQHKLKH--------- 162
                           ||: |      ||   | :| | |   | |::|:  |
Db        325 YVCEECGKAFKYSRILTTHKRIHTGEKPYKCNKCGKAFIASSTLSRHEFIHMGKKHYKCE 384

Qy        163 ------------------HGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSH---PKLKCPK 201
                             | |   |:|:   | | |:  | | | |:||     || :
Db        385 ECGKAFIWSSVLTRHKRVHTGEKPYKCEE--CGKAFKYSSTLSSH-KRSHTGEKPYKCEE 441

Qy        202 CGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGNIPDDL 260
              |||  |    || | : |   | | : |: |  || | : : | :|  | | |  |
Db        442 CGKAFVASSTLSKHEIIH---TGKKPYKCEEC--GKAFNQSSSLTKHKKI-HTGEKP--- 492

Qy        261 LKETEVKKLENLLDQGSKLNNLHELET--EKLKVEEDEEDEEDS---LDEKRSDVRSDSM 315
              |  |  |  |  | | |    :: |  :  | ||  :    |   :  |:  |
Db        493 YKCEECGKAFN---QSSSLTKHKKIHTGEKPYKCEECGKAFNQSSTLIKHKKIHTREKPY 549

Qy        316 SAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
                :  |:|  |   :  |::   |  |   | : |:   |  | | | |
Db        550 KCEECGKAF--HLSTHLTTHKILHTGEKPYRC--RECGKAFNHSATLSSHKKIH 599


RESULT 15
ABG16953
ID   ABG16953 standard; protein; 719 AA.
XX
AC   ABG16953;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #16944.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US008631.
XX
PR   31-MAR-2000; 2000US-00540217.
PR   23-AUG-2000; 2000US-00649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS81140.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity.
XX
PS   Claim 20; SEQ ID NO 47312; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC   sequences. (I) is useful as hybridisation probes, polymerase chain
CC   reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC   and in recombinant production of (II). The polynucleotides are also used
CC   in diagnostics as expressed sequence tags for identifying expressed
CC   genes. (I) is useful in gene therapy techniques to restore normal
CC   activity of (II) or to treat disease states involving (II). (II) is
CC   useful for generating antibodies against it, detecting or quantitating a
CC   polypeptide in tissue, as molecular weight markers and as a food
CC   supplement. (II) and its binding partners are useful in medical imaging
CC   of sites expressing (II). (I) and (II) are useful for treating disorders
CC   involving aberrant protein expression or biological activity. The
CC   polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC   amino acid sequences of the invention. Note: The sequence data for this
CC   patent did not appear in the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
```

SQ Sequence 719 AA;

Query Match 17.7%; Score 395; DB 4; Length 719;
Best Local Similarity 30.0%; Pred. No. 2.2e-22;
Matches 121; Conservative 51; Mismatches 174; Indels 58; Gaps 18;

```
Qy          3 ESDETKSISSLISSSS--SSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVD 60
              | |:  |  |:::     :  | | |  | | ||: | | | :| |:|: ::||||  :
Db        105 ECDKAFSKFSILTKHKVIHTGEKHYKC--EECGKAFTRSSSLIEHKRSHAGEKPYKC--E 160

Qy         61 DCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQE 120
              :| ||| : | |  |   |: :||: |  |||  |   :|  |:  ||    |  | | :
Db        161 ECGKAFSKASTLTAHKTIHAGEKPYKCEECGKAFNRSSNLMEHKRIHTGEKPCKCEECGK 220

Qy        121 AFYKHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCD------- 172
              ||    :| :| ::   ||   |::| | |: || | :||  | |  | |:|:
Db        221 AFGNFSTLTKHKVIHTGEKPYKCEECGKAFSWPSSLTEHKRIHAGDKP-YKCEECGKTFK 279

Qy        173 ----------HPG--------CFKNFQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKG 211
                       | |        | | | |:| |  |  |        || :|||
Db        280 WSSTLTKHKIIHTGEKPYKCEECGKAFTTFSSLTKHKVIHTGEKHYKCEECGKVFSWSSS 339

Qy        212 LSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLE 270
              |::|   |      |:: |: |  || |   : | ||  | | |  |   | |  |  |
Db        340 LTTHKAIHAGE---KLYKCEEC--GKAFKWSSRLSEHKRI-HTGEKP---YKCEECGKAF 390

Qy        271 NLLDQGSKLNNLHELETEK--LKVEEDEEDEEDS---LDEKRSDVRSDSMSAQRSIKSFT 325
              : :   | ||   ::   |   | ||  :|   |      ||          :   |:|
Db        391 SWV---SVLNKHKKIHAGKKFYKCEECGKDFNQSSHLTTHKRIHTGEKPYKCEECGKAF- 446

Qy        326 ASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
              |   : :  | |     |   |    | : |:|  :|  | : |
Db        447 -SKASTLTAHKTIHAGEKPYKC--EECGKAFNRSSNLMEHKRIH 487
```

Search completed: May 5, 2004, 15:21:25
Job time : 60 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:18:46 ; Search time 44 Seconds
(without alignments)
2954.398 Million cell updates/sec

Title: US-09-831-804-3
Perfect score: 2229
Sequence: 1 MSESDETKSISSLISSSSSS.....PLVKKARMDLLPNETSVISR 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	410.5	18.4	634	11	Q8BI85	Q8bi85 mus musculu
2	400	17.9	404	4	Q8N7Q3	Q8n7q3 homo sapien
3	397.5	17.8	395	4	O43693	O43693 homo sapien
4	392.5	17.6	542	4	Q8IYN0	Q8iyn0 homo sapien
5	392	17.6	555	4	Q96N38	Q96n38 homo sapien
6	392	17.6	644	4	Q8ND40	Q8nd40 homo sapien
7	392	17.6	665	4	Q86W65	Q86w65 homo sapien
8	391	17.5	592	4	Q8N8Q4	Q8n8q4 homo sapien
9	390	17.5	540	4	Q8N211	Q8n211 homo sapien
10	390	17.5	576	4	Q96N22	Q96n22 homo sapien
11	388.5	17.4	501	5	Q9VG72	Q9vg72 drosophila
12	388	17.4	568	4	Q8TD23	Q8td23 homo sapien
13	387	17.4	559	11	Q80VH2	Q80vh2 mus musculu
14	385.5	17.3	297	11	Q8BJ46	Q8bj46 mus musculu
15	382	17.1	511	11	Q8VEC1	Q8vec1 mus musculu
16	381	17.1	493	11	Q921H6	Q921h6 mus musculu
17	379.5	17.0	710	11	Q91VP4	Q91vp4 mus musculu
18	379	17.0	600	11	Q8CE00	Q8ce00 mus musculu
19	379	17.0	766	11	Q8C4J2	Q8c4j2 mus musculu
20	378.5	17.0	623	11	Q62510	Q62510 mus musculu
21	378.5	17.0	914	11	Q8C827	Q8c827 mus musculu
22	378	17.0	393	11	Q923E2	Q923e2 mus musculu
23	378	17.0	452	11	Q8BQC2	Q8bqc2 mus musculu
24	378	17.0	452	11	Q8BPP0	Q8bpp0 mus musculu
25	378	17.0	613	4	Q9H7R5	Q9h7r5 homo sapien
26	376	16.9	751	4	Q8TBA9	Q8tba9 homo sapien
27	374.5	16.8	710	11	Q61898	Q61898 mus musculu
28	374	16.8	578	11	Q8BLA2	Q8bla2 mus musculu
29	374	16.8	578	11	Q7TMN8	Q7tmn8 mus musculu
30	374	16.8	641	4	Q86T91	Q86t91 homo sapien
31	374	16.8	641	4	Q86T29	Q86t29 homo sapien
32	374	16.8	650	11	Q91WF9	Q91wf9 mus musculu
33	373.5	16.8	517	4	Q8N492	Q8n492 homo sapien
34	373	16.7	400	11	Q8VHT7	Q8vht7 mus musculu
35	373	16.7	747	4	Q86XU8	Q86xu8 homo sapien
36	372.5	16.7	586	4	Q8NB35	Q8nb35 homo sapien
37	372	16.7	637	4	Q96IR2	Q96ir2 homo sapien
38	372	16.7	737	11	Q7TSH9	Q7tsh9 mus musculu
39	369.5	16.6	464	4	Q8N8T4	Q8n8t4 homo sapien
40	369	16.6	531	4	Q8NE34	Q8ne34 homo sapien
41	369	16.6	650	4	Q86XL7	Q86xl7 homo sapien
42	368	16.5	488	4	Q9NV05	Q9nv05 homo sapien
43	368	16.5	488	4	Q96K00	Q96k00 homo sapien
44	367.5	16.5	462	11	Q8K286	Q8k286 mus musculu
45	366.5	16.4	393	4	Q9H6Z6	Q9h6z6 homo sapien

ALIGNMENTS

RESULT 1
Q8BI85
ID Q8BI85 PRELIMINARY; PRT; 634 AA.
AC Q8BI85;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to DNA-binding protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cortex;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK043955; BAC31714.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR001909; KRAB.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF01352; KRAB; 1.
DR Pfam; PF00096; zf-C2H2; 16.
DR PRINTS; PR00048; ZINCFINGER.
DR SMART; SM00349; KRAB; 1.
DR SMART; SM00355; ZnF_C2H2; 16.
DR PROSITE; PS50805; KRAB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 16.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 16.
SQ SEQUENCE 634 AA; 73668 MW; AA7DEE33D87508BA CRC64;

Query Match 18.4%; Score 410.5; DB 11; Length 634;

Best Local Similarity 32.8%; Pred. No. 2e-19;
Matches 117; Conservative 54; Mismatches 129; Indels 57; Gaps 19;

```
Qy          23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
               | | |  : | ||: : | |::||  |:  :|||| ::|||:|  || | ||   |: |
Db         323 KPYKC--KECGKAFTQSSSLKKHLNLHTGKKPYKC--EECDKSFTEKSTLTTHKRIHTGK 378

Qy          83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLR-HHILSVHEKT 139
               ||: |::||| ||  :|| |:  ||  | :||  : | :||    ||: |  |   ||
Db         379 KPYKCNICGKSFNSCTNLKTHQTLHTGEKPYKC--KECGKAFPYMSSLKIHQNLHPGEKP 436

Qy         140 LTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPKL 197
               ||:|:| ||  | |  |:: | |  | |:|:   | |:|   : |: |  :
Db         437 YKCKECDKSFTLKSSLRTHQIIHTGEKP-YKCNI--CGKSFNQCTNLKTHQRLHTGEKPY 493

Qy         198 KCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDGNIP 257
               || :|||       | ||   |      | : |  ||   | :|: |::|  | | |
Db         494 KCKECGKSFHYWSSLKSHQNLHSGE---KPYKCKECD-KSFTEKSTLIKHQRI-HTG--- 545

Qy         258 DDLLKETEVKKLE--NLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDSM 315
                        |||  |: |:          |:| | |    |:
Db         546 ---------KKLYKCNICDK--SFTWCASLKTHK-KFHTGEK----------------PY 577

Qy         316 SAQRSIKSF--TASLEGSKSVSKLISNSGKKI-NCPKNNCDRMFSREYDLRRHLKWH 369
                  ::  |||   ::|:  : : :   ::|:|:  |  |:||| :|     ||| | |
Db         578 KCRKCGKSFPQLSTLKRHQKIHEDKGHTGEKLYKC--NDCDRSYSYHSSFRRHQKIH 632
```

RESULT 2
Q8N7Q3
ID Q8N7Q3 PRELIMINARY; PRT; 404 AA.
AC Q8N7Q3;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein PLJ40479.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Suzuki O., Sasaki N., Aotsuka S., Shoji T., Ichihara T., Shiohata N.,
RA Matsumoto K., Hirano M., Sano S., Nomura R., Yoshikawa Y.,
RA Matsumura Y., Moriya S., Chiba E., Momiyama H., Onogawa S.,
RA Kaeriyama S., Satoh N., Matsunawa H., Takahashi E., Kataoka R.,
RA Kuga N., Kuroda A., Satoh I., Kamata K., Takami S., Terashima Y.,
RA Watanabe M., Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A.,
RA Ishii S., Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,
RA Kimura K., Yamashita H., Matsuo K., Nakamura Y., Sekine M.,
RA Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K.,
RA Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,
RA Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AK097798; BAC05174.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF00096; zf-C2H2; 14.
DR PRINTS; PR00048; ZINCFINGER.
DR SMART; SM00355; ZnF_C2H2; 14.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 14.
KW Hypothetical protein; Metal-binding; Nuclear protein; Zinc;
KW Zinc-finger.
SQ SEQUENCE 404 AA; 45923 MW; B1B35F62F5DF0007 CRC64;

Query Match 17.9%; Score 400; DB 4; Length 404;
Best Local Similarity 33.9%; Pred. No. 6.3e-19;
Matches 121; Conservative 39; Mismatches 163; Indels 34; Gaps 19;

```
Qy          23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
               | | |  : | ||: : | |::||  |:  :|||| ::|||:|  || | ||   |: |
Db         39 KPYKC--EECGKAFNRSSILTKHKIIHTGEKPYKC--EECGKGFSSVSTLNTHKAIHAEE 94

Qy          83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHHILSVH--EK 138
               ||: |  ||| ||  |  |:  ||  | :||  | | :||    ||  |  :|  ||
Db         95 KPYKCEECGKASNSSSKLMEHKRIHTGEKPYKC--EECGKAFSWSSSLTEH-KRIHAGEK 151

Qy         139 TLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPK 196
                  |::| | | | | | :||: | |  | |:|:  || | |  | | |  |
Db         152 PYKCEECGKAFNRSSILTKHKIIHTGEKP-YKCE--GCGKAFSKVSTLNTHKAIHAEEKP 208

Qy         197 LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGN 255
               || :|||       |  |  |   |  | : |: |  || |:  : | ||  | | |
Db         209 YKCEECGKASNSSSKLMEHKRIH---TGEKPYKCEEC--GKAFSWSSSLTEHKRI-HAGE 262

Qy         256 IPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEE-DEEDEEDSLDEKRSDVRSDS 314
               |    | |  |        :|  :|  | :  | ||  :      |: |   : :
Db         263 KP---YKCEECGKAFTWSSSFTKHKRIHAAE-KPYKCEECGKGFSTFSILTKHKIIHTGE 318

Qy         315 --MSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
                   :  |:|  |     :  |:|    |  |    | : |||   | || : |
Db         319 KRYKCEECGKAF--SWSSILTEHKIIHTGEKPYKC--EECGKAFSRSSSLTRHKRIH 371
```

RESULT 3
O43693
ID O43693 PRELIMINARY; PRT; 395 AA.
AC O43693;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Zinc-finger protein (Fragment).
GN ZFS-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testicular tumor;
RX MEDLINE=98069676; PubMed=9406578;
RA Ogawa T., Poncelet D., Kinoshita Y., Noce T., Takeda M., Kawamoto K.,
RA Udagawa K., Lecocq P., Marine J., Martial J., Hosaka M.;
RT "Enhanced expression in seminoma of human zinc finger genes located on
RT chromosome 19.";
RL Cancer Genet. Cytogenet. 100:36-42(1998).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; D70831; BAA24050.1; -.
DR HSSP; P08048; 7ZNF.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0008270; F:zinc ion binding; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF00096; zf-C2H2; 13.
DR PRINTS; PR00048; ZINCFINGER.
DR ProDom; PD000003; Znf_C2H2; 10.
DR SMART; SM00355; ZnF_C2H2; 13.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Metal-binding; Nuclear protein; Zinc; Zinc-finger.
FT NON_TER 1 1
SQ SEQUENCE 395 AA; 45291 MW; D4E06B54DCE0BF9B CRC64;

Query Match 17.8%; Score 397.5; DB 4; Length 395;
Best Local Similarity 35.9%; Pred. No. 9e-19;
Matches 97; Conservative 35; Mismatches 87; Indels 51; Gaps 14;

```
Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
Db        133 KPYKC--KECGKAFNRSSTLTTHRKIHTGEKPYKC--EECGKALTQSSHLTTHKIIHTGE 188

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAF----------------- 122
Db        189 KPYKCKKCGKAFNQSAHLTTHEVIHTGEKPYKC--EKCGKAFNHFSHLTTHKIIHTGEKP 246

Qy        123 ---------YKHQS--LRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQC 171
Db        247 YKCKECGKAFKHSSTLTKHKIIHTGEKPYKCKECGKAFNQSSKLTEHK-KIHTGEKPYEC 305

Qy        172 DHPGCFKNFQTWSVLQFHIKQSHPK---LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIW 228
Db        306 EE--CGKAFNQSSNLTRH-KKSHTEEKPYKCEECGKALNRPSTLTIHKIIH---TGEKPY 359

Qy        229 TCDYCDVGK-FAKKNELVEHYNIFHDGNIP 257
Db        360 KCEEC--GKAFNQSSKLTKHKKI-HTGEKP 386
```

RESULT 4
Q8IYN0
ID   Q8IYN0   PRELIMINARY;   PRT;   542 AA.
AC   Q8IYN0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to zinc finger protein 208.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Strausberg R.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC035579; AAH35579.1; -.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 11.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 11.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
SQ   SEQUENCE   542 AA; 62756 MW; 395709FE7E094C9B CRC64;

Query Match 17.6%; Score 392.5; DB 4; Length 542;
Best Local Similarity 30.0%; Pred. No. 2.7e-18;
Matches 115; Conservative 39; Mismatches 132; Indels 97; Gaps 19;

```
Qy          8 KSISSLISSSSSSR----PKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCD 63
Db        212 KSFCMLLHLTQHKRFHITENSYQC--KDCGKAFNWFSTLTTHRRIHTGEKPYKC--EECG 267

Qy         64 KAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEA 121
Db        268 KAFNRSSHLTTHKIIHTGEKPYRCEECGKAFNRSSHLTTHKRIHTGVKPYKCT--ECGKA 325

Qy        122 FYKHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNF 180
Db        326 FNRSSHLTTHRIIHTGEKPYKCEECGKAFNQSSTLTTHKITHAGEKP-YKCEE--CGKAF 382

Qy        181 QTWSVLQFHIKQSHPK---LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK 237
Db        383 YRFSYLIKH-KTSHTGEKFYKCEECGKGFNWSSALTKHKRIH---TGEKPYKCEEC--GK 436

Qy        238 -FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDE 296
Db        437 AFNESSNLTTH-KMIHTGEKP---YKCDECGKAFN------------------------ 467

Qy        297 EDEEDSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMF 356
Db        468 --------------RSSQLTAH-----------------KMIHTGEKPYKC--EECGKAF 494

Qy        357 SREYDLRRH---------LKWHD 370
Db        495 NRSSTLTKHKITHTGEKSYKWEE 517
```

RESULT 5
Q96N38
ID   Q96N38   PRELIMINARY;   PRT;   555 AA.
AC   Q96N38;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ31444.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA   Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA   Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA   Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA   Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR   EMBL; AK056006; BAB71072.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF00096; zf-C2H2; 13.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 13.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 14.
KW   Hypothetical protein; Metal-binding; Nuclear protein; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   555 AA; 64025 MW; 75C246D4820FB920 CRC64;

Query Match 17.6%; Score 392; DB 4; Length 555;
Best Local Similarity 32.5%; Pred. No. 3e-18;
Matches 116; Conservative 45; Mismatches 162; Indels 34; Gaps 17;

```
Qy         21 RPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHS 80
Db        169 RENSYQC--EECDKVFKRFSTLTRHKRVHTGEKPFKC--EECGKAFKHSSTLTTHKMIHT 224

Qy         81 EKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVHE 137
Db        225 GEKPYRCEECGKAFYHSSHLTTHKVIHTGEKPFKC--EECGKAFNHPSALTTHKFIHVKE 282

Qy        138 KTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHP 195
```

```
Db        283 KPYKCEECDKAFNRFSYLTKHKI-IHSGEKSYKCEQ--CGKGFNWSSTLTKHRRIHTGEK 339

Qy        196 KLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDGN 255

Db        340 PYKCEECGKAFNVSSHLTTHKMIH---TGEKPYKCEEC--GKAFNHSSKLTIHKIIHTGE 394

Qy        256 IPDDLLKETEVKKLENLLDQGSKLNNLHELET-EKL-KVEEDEEDEEDSLD---EKRSDV 310

Db        395 KP---YKCEECGKAFN---QSSNLTKHKIIHTGEKLYKCEECGKAFNRSSNLTTHKRIHT 448

Qy        311 RSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLK 367

Db        449 GEKPYKCEECGKAFNRSSNLTK--HNIIHTGEKSYKC--EECGKAFNQSSTLTKHRK 501


RESULT 6
Q8ND40
ID   Q8ND40      PRELIMINARY;      PRT;   644 AA.
AC   Q8ND40;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein (Fragment).
GN   DKFZP547O168.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RA   Bloecker H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR   EMBL; AL834415; CAD39077.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 13.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 13.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 14.
KW   Hypothetical protein; Metal-binding; Nuclear protein; Zinc;
KW   Zinc-finger.
FT   NON_TER       1      1
SQ   SEQUENCE   644 AA;  74012 MW;  A3CBF2437279B863 CRC64;

  Query Match             17.6%;  Score 392;  DB 4;  Length 644;
  Best Local Similarity   32.5%;  Pred. No. 3.5e-18;
  Matches  116;  Conservative   45;  Mismatches  162;  Indels   34;  Gaps   17;

Qy         21 RPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHS 80

Db        289 RENSYQC--EECDKVFKRFSTLTRHKRVHTGEKPFKC--EECGKAFKHSSTLTTHKMIHT 344

Qy         81 EKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVHE 137

Db        345 GEKPYRCEECGKAFYHSSHLTTHKVIHTGEKPFKC--EECGKAFNHPSALTTHKFIHVKE 402

Qy        138 KTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHP 195

Db        403 KPYKCEECDKAFNRFSYLTKHKI-IHSGEKSYKCEQ--CGKGFNWSSTLTKHRRIHTGEK 459

Qy        196 KLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDGN 255

Db        460 PYKCEECGKAFNVSSHLTTHKMIH---TGEKPYKCEEC--GKAFNHSSKLTIHKIIHTGE 514

Qy        256 IPDDLLKETEVKKLENLLDQGSKLNNLHELET-EKL-KVEEDEEDEEDSLD---EKRSDV 310

Db        515 KP---YKCEECGKAFN---QSSNLTKHKIIHTGEKLYKCEECGKAFNRSSNLTTHKRIHT 568

Qy        311 RSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLK 367

Db        569 GEKPYKCEECGKAFNRSSNLTK--HNIIHTGEKSYKC--EECGKAFNQSSTLTKHRK 621


RESULT 7
Q86W65
ID   Q86W65      PRELIMINARY;      PRT;   665 AA.
AC   Q86W65;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to zinc finger protein 91 (HPF7, HTF10) (Fragment).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Strausberg R.;
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC050468; AAH50468.1; -.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 13.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 11.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 13.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 14.
FT   NON_TER       1      1
SQ   SEQUENCE   665 AA;  76467 MW;  1ABDA1E4E00249DF CRC64;

  Query Match             17.6%;  Score 392;  DB 4;  Length 665;
  Best Local Similarity   32.5%;  Pred. No. 3.7e-18;
  Matches  116;  Conservative   45;  Mismatches  162;  Indels   34;  Gaps   17;

Qy         21 RPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHS 80

Db        279 RENSYQC--EECDKVFKRFSTLTRHKRVHTGEKPFKC--EECGKAFKHSSTLTTHKMIHT 334

Qy         81 EKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVHE 137

Db        335 GEKPYRCEECGKAFYHSSHLTTHKVIHTGEKPFKC--EECGKAFNHPSALTTHKFIHVKE 392

Qy        138 KTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHP 195

Db        393 KPYKCEECDKAFNRFSYLTKHKI-IHSGEKSYKCEQ--CGKGFNWSSTLTKHRRIHTGEK 449

Qy        196 KLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDGN 255

Db        450 PYKCEECGKAFNVSSHLTTHKMIH---TGEKPYKCEEC--GKAFNHSSKLTIHKIIHTGE 504

Qy        256 IPDDLLKETEVKKLENLLDQGSKLNNLHELET-EKL-KVEEDEEDEEDSLD---EKRSDV 310

Db        505 KP---YKCEECGKAFN---QSSNLTKHKIIHTGEKLYKCEECGKAFNRSSNLTTHKRIHT 558

Qy        311 RSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLK 367

Db        559 GEKPYKCEECGKAFNRSSNLTK--HNIIHTGEKSYKC--EECGKAFNQSSTLTKHRK 611
```

```
Db        559 GEKPYKCEECGKAFNRSSNLTK--HNIIHTGEKSYKC--EECGKAFNQSSTLTKHRK 611


RESULT 8
Q8N8Q4
ID   Q8N8Q4      PRELIMINARY;      PRT;   592 AA.
AC   Q8N8Q4;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ39023.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA   Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA   Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA   Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR   EMBL; AK096342; BAC04764.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 15.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 5.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 16.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 15.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 16.
DR   PROSITE; PS00142; ZINC_PROTEASE; 4.
KW   Hypothetical protein; Metal-binding; Nuclear protein; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   592 AA;  67920 MW;  D116FB9366EB4BDE CRC64;

  Query Match             17.5%;  Score 391;  DB 4;  Length 592;
  Best Local Similarity   31.1%;  Pred. No. 3.8e-18;
  Matches  120;  Conservative   41;  Mismatches  161;  Indels   64;  Gaps   19;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
              | |||  | | ||:   | |  | | |: ::|||| | |||||   | |  | : |: |
Db        199 KPYIC--EECGKAFKYSSALNTHKRIHTGEKPYKC--DKCDKAFIASSTLSKHEIIHTGK 254

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHH--------- 131
              ||: |  |||  |    | :|:  ||  | :||  | | :|| :  :| |
Db        255 KPYKCEECGKAFNQSSTLTKHKKIHTGEKPYKC--EECGKAFNQSSTLTKHKKIHTGEKP 312

Qy        132 --------------ILSVH------EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQC 171
                           ||: |      ||   | :| | |   | |::|:  |  |   |:|
Db        313 YVCEECGKAFKYSRILTTHKRIHTEEKPYKCNKCGKAFIASSTLSRHEFIHM-GKKHYKC 371

Qy        172 DHPGCFKNFQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWT 229
              :  | | |  |||  |  :        || :||| |    || | : |  | | :
Db        372 EE--CGKAFIWSSVLTRHKRVHTGEKPYKCEECGKAFVASSTLSKHEIIH---TGKKPYK 426

Qy        230 CDYCDVGK-FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELET- 287
              |: |  || | : : | :|  | | |  |    |  |  |  |   | | |    :: |
Db        427 CEEC--GKAFNQSSSLTKHKKI-HTGEKP---YKCEECGKAFN---QSSSLTKHKKIHTG 477

Qy        288 -EKLKVEEDEEDEEDS---LDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGK 343
               :  | ||  :    |  :  |:   |       :   |:|   |    :  |::    |
Db        478 EKPYKCEECGKAFNQSSTLIKHKKIHTREKPYKCEECGKAF--HLSTHLTTHKILHTGEK 535

Qy        344 KINCPKNNCDRMFSREYDLRRHLKWH 369
                 |    | : |:    |  | | |
Db        536 PYRC--RECGKAFNHSATLSSHKKIH 559


RESULT 9
Q8N211
ID   Q8N211      PRELIMINARY;      PRT;   540 AA.
AC   Q8N211;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein FLJ36350.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Thymus;
RA   Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,
RA   Arita M., Musashino K., Yuuki H., Hara H., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA   Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA   Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Murakawa K., Kanehori K., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR   EMBL; AK093669; BAC04216.1; -.
DR   PIR; F42075; F42075.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 12.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 12.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
KW   Hypothetical protein; Metal-binding; Nuclear protein; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   540 AA;  62985 MW;  84D7D2EC15A7002C CRC64;

  Query Match             17.5%;  Score 390;  DB 4;  Length 540;
  Best Local Similarity   30.0%;  Pred. No. 4e-18;
  Matches  107;  Conservative   61;  Mismatches  155;  Indels   34;  Gaps   15;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
              | | |  : ||: :|: | | :: : :: ::|||| ::| ||| : ||| || : |: :
Db        198 KLYKC--QECDRTFNQFSNLTEYKKDYARKPYKC--EECGKAFNQSSHLTTHKIIHTGE 253

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQEAFYKHQSL-RHHILSVHEKTLT 141
              ||: |  |||  |   :|  |: ||      | | :|| :  :|  |  :  ||
Db        254 KPYKCEECGKAFNQFSNLTTHKKIHTGEQPYICEECGKAFTQSSTLTTHKRIHTGEKPYK 313

Qy        142 CKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPKLKC 199
              |::| | | | ||| :||  | |  | |:|:   | | |   | |  |  |        ||
```

```
Db        314 CEECGKAFNRSSKLTEHKNIHTGEQP-YKCEE--CGKAFNRSSNLTEHRKIHTEEKPYKC 370

Qy        200 PKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGNIPD 258

Db        371 KECGKAFKHSSALTTHKRIH---TGEKPYKCEEC--GKAFNRSSKLTEHKKL-HTGK--- 421

Qy        259 DLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDE-----EDEEDSLDEKRSDVRSD 313

Db        422 ---KPYKCEECGKAFIQSSKLTEHKKIHSGEIPYKCEECGKAFKHSSSLTTHKRIHTGEK 478

Qy        314 SMSAQRSIKSFTASLEGSKSVSKLISNSGKK-INCPKNNCDRMFSREYDLRRHLKWH 369

Db        479 PYKCEECGKAFSRS---SKLTEHKIIHTGEKPYKCER--CDKAFNQSANLTKHKKIH 530


RESULT 10
Q96N22
ID   Q96N22      PRELIMINARY;      PRT;   576 AA.
AC   Q96N22;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Hypothetical protein FLJ31526.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA   Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA   Katsuta N., Sato K., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Murakawa K.,
RA   Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR   EMBL; AK056088; BAB71090.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF00096; zf-C2H2; 15.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 2.
DR   SMART; SM00355; ZnF_C2H2; 16.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 10.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 16.
KW   Hypothetical protein; Metal-binding; Nuclear protein; Zinc;
KW   Zinc-finger.
SQ   SEQUENCE   576 AA;  66620 MW;  4396672D34BF99CD CRC64;

  Query Match           17.5%;  Score 390;  DB 4;  Length 576;
  Best Local Similarity   28.6%;  Pred. No. 4.3e-18;
  Matches  119;  Conservative   57;  Mismatches  168;  Indels   72;  Gaps   17;

Qy          3 ESDETKSISSLISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDC 62

Db        217 ECDKAFNQSSTLTTHKIIHTREKLNEYKECGKAFNQSSHLTRHKIIHTGEKPYKC--EEC 274

Qy         63 DKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQE 120

Db        275 GKAFNQSSHLTRHKIIHTGEKPYKCEECGKAFRQSSHLTTHKIIHTGEKPYKC--EECGK 332

Qy        121 AFYKHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKN 179

Db        333 AFNKSSHLTRHKSIHTGEKPYQCEKCGKASNQSSNLTEHKNIHTEEKP-YKCEE--CGKA 389

Qy        180 FQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK 237

Db        390 FNQFSNLTTHKRIHTGEKPYKCEECGKAFNQSSILTTHKRIH---TGEKSYKCEEC--GK 444

Qy        238 -FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDE 296

Db        445 AFYRSSKLTEHKKI-HTGEKP-----------YTCEECGKAFNHSSHLATHKV-IHTGE 490

Qy        297 EDEEDSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMF 356

Db        491 KPYQ----------------CEECGKAFNQSSHLTR--HKRIHTGEKPYQCEK--CGKAF 530

Qy        357 SREYDLRRHLKWHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNETSVISR 412

Db        531 NQSSNLTGHKKIH-------------------TGEKLYKPKRCNSDFENTSKFSK 566


RESULT 11
Q9VG72
ID   Q9VG72      PRELIMINARY;      PRT;   501 AA.
AC   Q9VG72;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   CG5245 protein.
GN   CG5245.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003696; AAF54813.1; -.
DR   HSSP; P08047; 1SP2.
DR   FlyBase; FBgn0038047; CG5245.
```

```
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 15.
DR   ProDom; PD000003; Znf_C2H2; 3.
DR   SMART; SM00355; ZnF_C2H2; 15.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 15.
KW   Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE   501 AA;  58790 MW;  BBAE04740D41C43F CRC64;

  Query Match           17.4%;  Score 388.5;  DB 5;  Length 501;
  Best Local Similarity   29.7%;  Pred. No. 4.6e-18;
  Matches  109;  Conservative   63;  Mismatches  142;  Indels   53;  Gaps   19;

Qy         21 RPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHS 80
              ||  |  |::  |  | : : |  |:|||   |: :||:|||    |  :| |||||: |: :||
Db        135 RPHK--CSH--CSKTFTQNSSLKQHLHEHTGERPFKCT--QCSTSFARKSHLQVHLRTHS 188

Qy         81 EKKPFHCSVCGKGVNSRQHLKRHEITH--TKSFKCTFENCQEAFYKHQSLRHHILSVHEK 138
              |::|| |: | |   :  ||: |  ||   : |||:  :| ::|     |: |:|:  |:
Db        189 EERPFECTHCEKAFKNNSHLQEHLRTHQEARPFKCS--HCSKSFKLRSILQKHLLTHAER 246

Qy        139 TLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK-- 196
              :  | || | | :   |  | |: | |   ::| |  | : |   | || |: :   |
Db        247 SFKCTQCPKTFLQNDSLQIH-LRVHAGEDPFKCPH--CSETFARNSRLQLHLLEHAGKEP 303

Qy        197 LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDGNI 256
              ||| :|      :     |:  |   |  : : |   |     | ||: |||| : | |
Db        304 LKCSQCSATFAMRSLYRVHVRLH---TRERQYKCAECS-KSFFKKSHLVEHQQV-HTGER 358

Qy        257 P-------DDLLKETEVKKLENLLDQ-GSKLNNLHELETE-----KLKVEEDEEDEEDSL 303
              |       |  | ::   ::|| | |:          |     :| |   |    ::
Db        359 PFKCTHCFKDFKCRTHLR--VHMLDHIGEKVPKCSYCSKEFKLSSQLLVHLQEHTGKNQF 416

Qy        304 DEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGK-KINCPKNNCDRMFSREYDL 362
              :                ||:| |    |     | :::|:    |  ::| ::|:|  :
Db        417 E------------CPHCSKSYTTS---STLHMHLRTHTGELPFKC--SHCSKLFARSAEH 459

Qy        363 RRHLKWH 369
              : ||: |
Db        460 QEHLRTH 466


RESULT 12
Q8TD23
ID   Q8TD23      PRELIMINARY;      PRT;   568 AA.
AC   Q8TD23;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   TRAF6-binding zinc finger protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21864204; PubMed=11751921;
RA   Shin J.N., Kim I., Lee J.S., Koh G.Y., Lee Z.H., Kim H.H.;
RT   "A Novel Zinc Finger Protein That Inhibits Osteoclastogenesis and the
RT   Function of Tumor Necrosis Factor Receptor-associated Factor 6.";
RL   J. Biol. Chem. 277:8346-8353(2002).
DR   EMBL; AY044432; AAK95822.1; -.
DR   PIR; F42075; F42075.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 13.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 13.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 9.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 14.
KW   Metal-binding; Zinc; Zinc-finger.
SQ   SEQUENCE   568 AA;  66222 MW;  3BBC819822B89940 CRC64;

  Query Match           17.4%;  Score 388;  DB 4;  Length 568;
  Best Local Similarity   28.5%;  Pred. No. 5.7e-18;
  Matches  105;  Conservative   62;  Mismatches  167;  Indels   34;  Gaps   13;

Qy         15 SSSSSSRPKKYIC----TYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKS 70
              ||  :   : | |       : ||: :|: | | :: : :: ::|||| ::| ||| : |
Db        212 SSKLTKHKRIYTCEKLYKCQECDRTFNQFSNLTEYKKDYAREKPYKC--EECGKAFNQSS 269

Qy         71 HLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQEAFYKHQSL-R 129
              || || : | : :||: |  |||  |   :|  |: ||        | | :|| :  :|
Db        270 HLTTHKIIHTGEKPYKCEECGKASNQFSNLTTHKKIHTGEQPYICEECGKAFTQSSTLTT 329

Qy        130 HHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH 189
              |  :   ||   |::| | | | ||| :||  | |  | |:|:   | | |   | |  |
Db        330 HKRIHTGEKPYKCEECGKAFNRSSKLTEHKNIHTGEQP-YKCEE--CGKAFNRSSNLTEH 386

Qy        190 --IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEH 247
                |     || :|||        |::|   |  | | : |: |  ||   ::  :
Db        387 RKIHTEEKPYKCKECGKAFKHSSALTTHKRIH---TGEKPYKCEEC--GKAFNRSSKLTE 441

Qy        248 YNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDE-----EDEEDS 302
              |   | |        |  : ::      | |||     :: : ::  : :|     :
Db        442 YKKLHTGK------KPYKCEECGKAFIQSSKLTEHKKIHSGEIPYKCEECGKAFKHSSSL 495

Qy        303 LDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKK-INCPKNNCDRMFSREYD 361
                ||           :  |:|: |   ||     | ::|:|   | :  ||: |::  :
Db        496 TTHKRIHTGEKPYKCEECSKAFSRS---SKLTEHKIIHTGEKPYECER--CDKAFNQSAN 550

Qy        362 LRRHLKWH 369
              | :| | |
Db        551 LTKHKKIH 558


RESULT 13
Q80VH2
ID   Q80VH2      PRELIMINARY;      PRT;   559 AA.
AC   Q80VH2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to zinc finger protein 97.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FVB/N; TISSUE=Breast tumor;
RA   Strausberg R.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC049144; AAH49144.1; -.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0003676; F:nucleic acid binding; IEA.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 15.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 15.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 15.
SQ   SEQUENCE   559 AA;  66355 MW;  F456D3FB56E86C39 CRC64;
```

Query Match 17.4%; Score 387; DB 11; Length 559;
Best Local Similarity 28.7%; Pred. No. 6.6e-18;
Matches 109; Conservative 55; Mismatches 136; Indels 80; Gaps 17;

```
Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
Db        187 KPYKCNQ--CDKAFSQHSTLQTHRRTHTGEKPYKC--DQCDKAFSEKCSLQTHRRTHTGE 242

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKC-----TFEN----------------- 117
Db        243 KPYKCNQCDKAFSQYSHLHIHRRTHTGEKPLKCNECDETFSNHSNLQTHRRIHTGEKPYK 302

Qy        118 ---CQEAFYKHQSLRHHILS-VHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDH 173
Db        303 CNQCDKAFSQHSTLQNHRRTHTGEKPFKCNQCDKAFSRHSTLQTHRRTHTGEKP-FKCNQ 361

Qy        174 PGCFKNFQTWSVLQFHIKQSH---PKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTC 230
Db        362 --CDKAFSQYSHLHIH-RRTHTGEKPFKCNQCNKAFSQYSHLHIHRRTH---TGEKPYKC 415

Qy        231 DYCDVGKFAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKL 290
Db        416 NQCD--KTFSNHSTLQTHRRTHTGE------KPYKCNQCDKAFSRHSTLQTHRRTHTGEK 467

Qy        291 KVEEDEEDEEDSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKI-NCPK 349
Db        468 PFKCNQCDK----------AFSQKCSLQKHIRIHT---------------GEKLYKC-- 499

Qy        350 NNCDRMFSREYDLRRHLKWH 369
Db        500 NECDKAFSQHSTLQTHRRTH 519
```

RESULT 14
Q8BJ46
ID Q8BJ46 PRELIMINARY; PRT; 297 AA.
AC Q8BJ46;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MSZF33 homolog.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Olfactory brain;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK032220; BAC27766.1; -.
DR InterPro; IPR007087; Znf_C2H2.
DR InterPro; IPR007086; Znf_C2H2_sub.
DR Pfam; PF00096; zf-C2H2; 8.
DR PRINTS; PR00048; ZINCFINGER.
DR SMART; SM00355; ZnF_C2H2; 8.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
SQ SEQUENCE 297 AA; 34666 MW; 2B94CC27CB2635EB CRC64;

Query Match 17.3%; Score 385.5; DB 11; Length 297;
Best Local Similarity 38.1%; Pred. No. 4.2e-18;
Matches 93; Conservative 36; Mismatches 82; Indels 33; Gaps 14;

```
Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
Db         69 KRYKCNQ--CDKAYSRHSILQIHKRTHSGEKPYEC--NQCGKAFTQHSHLKIHMVTHTGE 124

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHHILSVHEKTL 140
Db        125 KPYKCNQCGKAFACHNKLQKHERTHTGEKPYKC--DQCNKAF-----VYESYLQVHKKTH 177

Qy        141 T------CKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSH 194
Db        178 TGEKPYKCNECGKAFARHSHLKVHKITHTGEKP-YKCNQ--CGKALAYHSTLQVH-QRTH 233

Qy        195 PKLK---CPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNI 250
Db        234 TGEKPYECEQCGKAFANQSYFQVHKRIH---TGEKPYKCDQC--GKAFVGSSDLKRHERV 288

Qy        251 FHDG 254
Db        289 -HTG 291
```

RESULT 15
Q8VEC1
ID Q8VEC1 PRELIMINARY; PRT; 511 AA.
AC Q8VEC1;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Similar to zinc finger protein 40.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC019219; AAH19219.1; -.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 12.
DR SMART; SM00355; ZnF_C2H2; 12.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 511 AA; 59582 MW; 11C75B5E38F2DB6F CRC64;

Query Match 17.1%; Score 382; DB 11; Length 511;
Best Local Similarity 31.1%; Pred. No. 1.3e-17;
Matches 110; Conservative 47; Mismatches 143; Indels 54; Gaps 17;

```
Qy         18 SSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIV 77
Db        138 TNNREKTYKCS--ECDKCFTKKCKLRQHQRIHTGEKPYKCS--ECDKCFTQKDRLIIHQR 193

Qy         78 SHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHHILSV 135
Db        194 VHTGEKPYKCSECDKGFTSKSRLNIHQRVHTGEKSYKCS--ECDKSFSQQGNLSIH-LRI 250

Qy        136 H--EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IK 191
Db        251 HTGEKPYKCSECDKCFTYKSGLRSHQI-IHTGEKAYKCSE--CGKCFTHKGDLRRHQSIH 307

Qy        192 QSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIF 251
Db        308 MGEKPYKCSECDKGFTQKCRLIIHQIIHTGKKGYKCSECDKC----FILKSDLSIHQRI- 362

Qy        252 HDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVR 311
Db        363 HTGE------KPYECSECGKGFTSKSRL-NIHQ------RVHTGEK-------------- 395

Qy        312 SDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRH 365
Db        396 --SYKCSECDKSFTQ--QGNLRIHLRIHTGEKPYKC--SECDKCFTYKSGLRSH 443
```

Search completed: May 5, 2004, 15:22:52
Job time : 46 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:15:26 ; Search time 17 Seconds
(without alignments)
1261.936 Million cell updates/sec

Title: US-09-831-804-3
Perfect score: 2229
Sequence: 1 MSESDETKSISSLISSSSSS.....PLVKKARMDLLPNETSVISR 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	837.5	37.6	429	1	TF3A_YEAST	P39933 saccharomyc
2	495	22.2	374	1	TF3A_SCHPO	Q9utl5 schizosacch
3	400	17.9	803	1	ZN43_HUMAN	P17038 homo sapien
4	398	17.9	574	1	Z492_HUMAN	Q9p255 homo sapien
5	397.5	17.8	423	1	TF3A_HUMAN	Q92664 homo sapien
6	396.5	17.8	595	1	ZN85_HUMAN	Q03923 homo sapien
7	395	17.7	616	1	ZN93_HUMAN	P35789 homo sapien
8	393.5	17.7	1191	1	ZN91_HUMAN	Q05481 homo sapien
9	387.5	17.4	839	1	Z347_HUMAN	Q96se7 homo sapien
10	385	17.3	576	1	Z431_HUMAN	Q8tf32 homo sapien
11	383	17.2	913	1	Z228_HUMAN	Q9uju3 homo sapien
12	376	16.9	570	1	Z430_HUMAN	Q9h8g1 homo sapien
13	376	16.9	751	1	Z184_HUMAN	Q99676 homo sapien
14	375.5	16.8	1350	1	XFIN_XENLA	P08045 xenopus lae
15	374	16.8	535	1	Z257_HUMAN	Q9y2q1 homo sapien
16	373	16.7	470	1	Z436_HUMAN	Q9c0f3 homo sapien
17	373	16.7	738	1	ZN84_HUMAN	P51523 homo sapien
18	373	16.7	821	1	ZN41_HUMAN	P51814 homo sapien
19	372	16.7	393	1	ZF46_MOUSE	Q03309 mus musculu
20	367	16.5	818	1	KR18_HUMAN	Q9hcg1 homo sapien
21	364	16.3	428	1	ZN83_HUMAN	P51522 homo sapien
22	363.5	16.3	898	1	ZO71_XENLA	P18751 xenopus lae
23	361	16.2	803	1	Z226_HUMAN	Q9nyt6 homo sapien
24	360.5	16.2	630	1	PRD5_HUMAN	Q9nqx1 homo sapien
25	360	16.2	588	1	Z429_HUMAN	Q86v71 homo sapien
26	359	16.1	1167	1	Z208_HUMAN	O43345 homo sapien
27	357.5	16.0	739	1	YD49_HUMAN	Q9p2j8 homo sapien
28	355	15.9	538	1	Z155_HUMAN	Q12901 homo sapien
29	354	15.9	498	1	Z271_HUMAN	Q14591 homo sapien
30	354	15.9	698	1	Z234_HUMAN	Q14588 homo sapien
31	353	15.8	604	1	Z305_HUMAN	O43309 homo sapien
32	351	15.7	474	1	Z141_HUMAN	Q15928 homo sapien
33	350.5	15.7	783	1	ZF25_HUMAN	Q9uii5 homo sapien
34	350	15.7	670	1	ZN16_HUMAN	P17020 homo sapien
35	350	15.7	706	1	Z225_HUMAN	Q9uk10 homo sapien
36	349	15.7	489	1	ZN71_HUMAN	Q9ngz8 homo sapien
37	348.5	15.6	439	1	ZO28_XENLA	P18747 xenopus lae
38	348.5	15.6	456	1	ZN25_HUMAN	P17030 homo sapien
39	348.5	15.6	606	1	Z214_HUMAN	Q9ul59 homo sapien
40	348.5	15.6	707	1	ZF60_MOUSE	P16374 mus musculu
41	348	15.6	617	1	Z221_HUMAN	Q9uk13 homo sapien
42	347	15.6	1029	1	Z197_HUMAN	O14709 homo sapien
43	346	15.5	623	1	Z255_HUMAN	Q9uid9 homo sapien
44	345.5	15.5	335	1	TF3A_RANPI	P34695 rana pipien
45	345.5	15.5	643	1	ZN23_HUMAN	P17027 homo sapien

ALIGNMENTS

RESULT 1
TF3A_YEAST
ID TF3A_YEAST STANDARD; PRT; 429 AA.
AC P39933;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor IIIA (TFIIIA).
GN TFC2 OR PZF1 OR TFIIIA OR YPR186C OR P9677.9.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92147684; PubMed=1737784;
RA Archambault J., Milne C.A., Schappert K.T., Baum B., Friesen J.D.,
RA Segall J.;
RT "The deduced sequence of the transcription factor TFIIIA from
RT Saccharomyces cerevisiae reveals extensive divergence from Xenopus
RT TFIIIA.";
RL J. Biol. Chem. 267:3282-3288(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92237295; PubMed=1570325;
RA Woychik N.A., Young R.A.;
RT "Genes encoding transcription factor IIIA and the RNA polymerase
RT common subunit RPB6 are divergently transcribed in Saccharomyces
RT cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:3999-4003(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
RA Hunicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
RA Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL Nature 387:103-105(1997).
CC -!- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
CC APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES, IS REQUIRED FOR
CC CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
CC BINDS THE TRANSCRIBED 5S RNA'S.

```
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M80611; AAB08014.1; -.
DR   EMBL; M90638; -; NOT_ANNOTATED_CDS.
DR   EMBL; U25841; AAB64615.1; -.
DR   PIR; S20050; S20050.
DR   GermOnline; 144451; -.
DR   TRANSFAC; T03530; -.
DR   SGD; S0006390; PZF1.
DR   GO; GO:0005667; C:transcription factor complex; IPI.
DR   GO; GO:0003709; F:RNA polymerase III transcription factor act. . .; IPI.
DR   GO; GO:0006384; P:transcription initiation from Pol III promoter; IPI.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 9.
DR   SMART; SM00355; ZnF_C2H2; 9.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 6.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   RNA-binding; Repeat; Nuclear protein.
FT   DOMAIN       24     43       SER-RICH.
FT   ZN_FING      49     74       C2H2-TYPE.
FT   ZN_FING      80    102       C2H2-TYPE.
FT   ZN_FING     108    130       C2H2-TYPE.
FT   ZN_FING     134    159       C2H2-TYPE.
FT   ZN_FING     163    186       C2H2-TYPE.
FT   ZN_FING     194    219       C2H2-TYPE.
FT   ZN_FING     222    244       C2H2-TYPE.
FT   ZN_FING     253    277       C2H2-TYPE.
FT   DOMAIN      321    327       ARG/LYS-RICH (BASIC).
FT   ZN_FING     365    389       C2H2-TYPE.
SQ   SEQUENCE   429 AA;  50027 MW;  209B1EDEA20422D9 CRC64;

  Query Match             37.6%;  Score 837.5;  DB 1;  Length 429;
  Best Local Similarity   43.7%;  Pred. No. 1.3e-42;
  Matches  174;  Conservative   62;  Mismatches  135;  Indels   27;  Gaps   11;

Qy          1 MSESDETKSISSLIS--SSSSSRPKKYICTYEGCDKAYNRPSLLEQH-LRTHSNDRPYKC 57
Db         23 ISRSESSESLNSLTSTRSSSSNRPKTYFCDYDGCDKAFTRPSILTEHQLSVHQGLRAFQC 82

Qy         58 TVDDCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFEN 117
Db         83 --DKCAKSFVKKSHLERHLYTHSDTKPFQCSYCGKGVTTRQQLKRHEVTHTKSFICPEEG 140

Qy        118 CQEAFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGG--SPAYQCDHPG 175
Db        141 CNLRFYKHPQLRAHILSVHLHKLTCPHCNKSFQRPYRLRNHISKHHDPEVENPYQCTFAG 200

Qy        176 CFKNFQTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDV 235
Db        201 CCKEFRIWSQLQSHIKNDHPKLKCPICSKPCVGENGLQMHMIIHDDSLVTKNWKCHICPD 260

Qy        236 GKFAKKNELVEHYNIFH-DGNIPDDL-LKETEVKKLENLLDDQGSKLNNLHELETEKLKVE 293
Db        261 MSFSRKHDLLTHYGSIHTEEDIPLELKYKISDIQQL--VQDHGVQLGN------SKHSNE 312

Qy        294 EDEEDEEDSLDEKRSDVRSDSMSAQRSIKSFTASLE-GSKSVSKLISNSGKKINCPKNNC 352
Db        313 QDEEKISNRLRKRRKLTENNNVEFLQNEVDLEKRLESGENGLNLLLNTVGRKYRCFYNNC 372

Qy        353 DRMFSREYDLRRHL---KWHDDNLQRIESFLNSIEKEE 387
Db        373 SRTFKTKEKYEKHIDKHKVHELKLKILQ------EKEE 404
```

```
RESULT 2
TF3A_SCHPO
ID   TF3A_SCHPO     STANDARD;      PRT;   374 AA.
AC   Q9UTL5;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Transcription factor IIIA (Factor A) (TFIIIA).
GN   SFC2 OR SPAC144.09C.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A., AND FUNCTION.
RX   MEDLINE=22082319; PubMed=12087160;
RA   Schulman D.B., Setzer D.R.;
RT   "Identification and characterization of transcription factor IIIA from
RT   Schizosaccharomyces pombe.";
RL   Nucleic Acids Res. 30:2772-2781(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
CC   -!- FUNCTION: Is required for correct transcription of 5S RNA genes by
CC       RNA polymerase III. Also binds the transcribed 5S RNA's. Initiates
CC       transcription of the 5S ribosomal RNA gene.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AY091590; AAM00046.1; -.
DR   EMBL; AL132675; CAB59689.1; -.
DR   PIR; T37676; T37676.
DR   HSSP; P07248; 1ARD.
DR   GeneDB_SPombe; SPAC144.09c; -.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00096; zf-C2H2; 10.
DR   ProDom; PD000003; Znf_C2H2; 1.
DR   SMART; SM00355; ZnF_C2H2; 9.
```

```
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 9.
KW   Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW   RNA-binding; Repeat; Nuclear protein.
FT   ZN_FING      23     47       C2H2-TYPE.
FT   ZN_FING      53     77       C2H2-TYPE.
FT   ZN_FING      83    107       C2H2-TYPE.
FT   ZN_FING     113    138       C2H2-TYPE.
FT   ZN_FING     144    169       C2H2-TYPE.
FT   ZN_FING     204    226       C2H2-TYPE.
FT   ZN_FING     236    261       C2H2-TYPE.
FT   ZN_FING     267    291       C2H2-TYPE (ATYPICAL).
FT   ZN_FING     349    374       C2H2-TYPE.
SQ   SEQUENCE   374 AA;  43851 MW;  7469C701FFF08FF6 CRC64;

  Query Match             22.2%;  Score 495;  DB 1;  Length 374;
  Best Local Similarity   32.8%;  Pred. No. 1.6e-22;
  Matches  123;  Conservative   58;  Mismatches  160;  Indels   34;  Gaps   13;

Qy         15 SSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLET 74
              |  :  |   | : | ||  |  | |:|||||||||||||||:||: |     | |||:|||||:
Db         13 SKNLRSAKKIFHCPYEECGKKYSRPSLLEQHLRTHSNERPFVCDYTGCSKAFYRKSHLKI 72

Qy         75 HIVSHSEKKPFHCSV--CGKGVNSRQHLKRHEITH--TKSFKCTFENCQEAFYKHQSLRH 130
              |    |:  ||| |     |       ::|||:||    |    | : ||:| | | | ||| ||
Db         73 HKRCHTNVKPFSCHYDGCDAQFYTQQHLERHIEVHRKPKPYACTWEGCDECFSKHQQLRS 132

Qy        131 HILSVHEKTL----TCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGC--FKNFQTWS 184
              || : |    |    | : |    |     ||   |  : |     :| | |   |   : |: ||
Db        133 HISACHTHLLPYPCTYQDCELRFATKQKLQNHVNRAHEKIISYSCPHESCVGHEGFEKWS 192

Qy        185 VLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDV----GKFAK 240
              || ||:::|    |  ||:       |  |:: |    | ::     :| :      | :
Db        193 QLQNHIREAHVP-SCSICGRQFKTAAHLRHHVVLH--QTTLEERKTYHCPMEGCKKSFTR 249

Qy        241 KNELVEHYNIFHDGNIP---DDLLKETEVK-KLENLLDQGSKLNNLHELETEKLKVEEDE 296
              : | :| :: |:||:     |    :   |  |:  |::|:     |:    :  :: |
Db        250 SSALKKHISVIHEGNMAFHCDSCGTKFGYKHMLQRHLERGT-CKKAHKPYINECGIKHDG 308

Qy        297 EDEEDSLDEKRSDVRSDSMS--AQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDR 354
              :     |:|  :: |: :|  |:: |    |    | |          :: :|    |:
Db        309 IEGVAIHDQKEKELSSNLVSDVAKKIINEVTG--HGYK--------EAREYSCSFPECNY 358

Qy        355 MFSREYDLRRHLKWH 369
              | | ||: |||  |
Db        359 RFKRLYDMHRHLNSH 373


RESULT 3
ZN43_HUMAN
ID   ZN43_HUMAN     STANDARD;      PRT;   803 AA.
AC   P17038; P28160; Q96DG1;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein 43 (Zinc protein HTF6) (Zinc finger protein
DE   KOX27).
GN   ZNF43 OR ZNF39 OR KOX27.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=T-cell;
RX   MEDLINE=91279444; PubMed=1711675;
RA   Lovering R., Trowsdale J.;
RT   "A gene encoding 22 highly related zinc fingers is expressed in
RT   lymphoid cell lines.";
RL   Nucleic Acids Res. 19:2921-2927(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymph;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   SEQUENCE OF 38-190 FROM N.A.
RX   MEDLINE=91219421; PubMed=2023909;
RA   Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
RA   Martial J.A.;
RT   "The evolutionarily conserved Kruppel-associated box domain defines a
RT   subfamily of eukaryotic multifingered proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
RN   [4]
RP   SEQUENCE OF 476-531 FROM N.A.
RC   TISSUE=Lymphoid;
RX   MEDLINE=91145339; PubMed=2288909;
RA   Thiesen H.-J.;
RT   "Multiple genes encoding zinc finger domains are expressed in human T
RT   cells.";
RL   New Biol. 2:363-374(1990).
CC   -!- FUNCTION: May be involved in transcriptional regulation.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: T AND B CELL LINES.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 1 KRAB domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X59244; CAA41932.1; -.
DR   EMBL; BC006528; AAH06528.1; -.
DR   EMBL; M61869; AAA58674.1; -.
DR   EMBL; X52358; CAA36584.1; -.
DR   PIR; S26823; S26823.
DR   HSSP; P08048; 7ZNF.
DR   TRANSFAC; T04986; -.
DR   Genew; HGNC:13109; ZNF43.
DR   MIM; 603972; -.
DR   GO; GO:0003677; F:DNA binding; TAS.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 21.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 16.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 22.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 19.
```

```
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 22.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat.
FT   DOMAIN        1     69       KRAB.
FT   ZN_FING     168    190       C2H2-TYPE.
FT   ZN_FING     196    218       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     224    246       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     252    274       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     280    302       C2H2-TYPE.
FT   ZN_FING     308    330       C2H2-TYPE.
FT   ZN_FING     336    358       C2H2-TYPE.
FT   ZN_FING     364    386       C2H2-TYPE.
FT   ZN_FING     392    414       C2H2-TYPE.
FT   ZN_FING     420    442       C2H2-TYPE.
FT   ZN_FING     448    470       C2H2-TYPE.
FT   ZN_FING     476    498       C2H2-TYPE.
FT   ZN_FING     504    526       C2H2-TYPE.
FT   ZN_FING     532    554       C2H2-TYPE.
FT   ZN_FING     560    582       C2H2-TYPE.
FT   ZN_FING     588    610       C2H2-TYPE.
FT   ZN_FING     616    638       C2H2-TYPE.
FT   ZN_FING     644    666       C2H2-TYPE.
FT   ZN_FING     672    694       C2H2-TYPE.
FT   ZN_FING     700    722       C2H2-TYPE.
FT   ZN_FING     728    750       C2H2-TYPE.
FT   ZN_FING     756    778       C2H2-TYPE.
FT   CONFLICT    181    181       P -> S (IN REF. 3).
FT   CONFLICT    712    712       P -> S (IN REF. 2).
SQ   SEQUENCE   803 AA;  93487 MW;  88C41B44DFDE2533 CRC64;

  Query Match             17.9%;  Score 400;  DB 1;  Length 803;
  Best Local Similarity   31.9%;  Pred. No. 1.4e-16;
  Matches  121;  Conservative   53;  Mismatches  171;  Indels   34;  Gaps   19;

Qy          2 SESDETKSISSLISSSSS--SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTV 59
              :|  |  | || ::      :  | | |  | | ||:   | | :|  ||: ::|||| 
Db        367 TECGEAFSRSSNLTKHKKIHTEKKPYKC--EECGKAFKWSSKLTEHKLTHTGEKPYKC-- 422

Qy         60 DDCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFEN 117
              ::| |||   | |  |   |: :||: | ||||  |   :|  |:  ||  | :||  | 
Db        423 EECGKAFNWPSTLTKHNRIHTGEKPYKCEVCGKAFNQFSNLTTHKRIHTAEKPYKC--EE 480

Qy        118 CQEAFYKHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGC 176
              | :|| :  :| :|  : : :|   |::| | |   ||| :||: | |  | |:|:  | 
Db        481 CGKAFSRSSNLTKHKKIHIEKKPYKCEECGKAFKWSSKLTEHKITHTGEKP-YKCEE--C 537

Qy        177 FKNFQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCD 234
              | |  :|:|  |  |        || :|||       |::|   |   | | : |: | 
Db        538 GKAFNHFSILTKHKRIHTGEKPYKCEECGKAFTQSSNLTTHKKIH---TGEKPYKCEEC- 593

Qy        235 VGK-FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVE 293
              || | : | | | | | |     | | | | |      :|  :| | : | | 
Db        594 -GKAFTQSSNLTTHKKI-HTGGKP---YKCEECGKAFNQFSTLTKHKIIH-TEEKPYKCE 647

Qy        294 E-DEEDEEDSLDEKRSDVRSDS--MSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKN 350
              |  :  :  |   |  : :       :  |:|   | : |  |:|    |  | | 
Db        648 ECGKAFKWSSTLTKHKIIHTGEKPYKCEECGKAF--KLSSTLSTHKIIHTGEKPYKCEK- 704

Qy        351 NCDRMFSREYDLRRHLKWH 369
              | : |:|  :|  | | | 
Db        705 -CGKAFNRPSNLIEHKKIH 722


RESULT 4
Z492_HUMAN
ID   Z492_HUMAN     STANDARD;      PRT;   574 AA.
AC   Q9P255;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Zinc finger protein 492 (Fragment).
GN   ZNF492 OR KIAA1473.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20277482; PubMed=10819331;
RA   Nagase T., Kikuno R., Ishikawa K.-I., Hirosawa M., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XVII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 7:143-150(2000).
CC   -!- FUNCTION: May function as a transcription factor.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 1 KRAB domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AB040906; BAA95997.1; -.
DR   HSSP; P07248; 1ARD.
DR   Genew; HGNC:23707; ZNF492.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 13.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 2.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 13.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
KW   Hypothetical protein; Transcription regulation; DNA-binding;
KW   Zinc-finger; Metal-binding; Nuclear protein; Repeat.
FT   NON_TER       1      1
FT   DOMAIN       15     86       KRAB.
FT   ZN_FING     184    206       C2H2-TYPE.
FT   ZN_FING     212    234       C2H2-TYPE.
FT   ZN_FING     240    262       C2H2-TYPE.
FT   ZN_FING     268    290       C2H2-TYPE.
FT   ZN_FING     296    318       C2H2-TYPE.
FT   ZN_FING     324    346       C2H2-TYPE.
FT   ZN_FING     352    374       C2H2-TYPE.
FT   ZN_FING     380    402       C2H2-TYPE.
FT   ZN_FING     408    430       C2H2-TYPE.
FT   ZN_FING     436    458       C2H2-TYPE.
FT   ZN_FING     464    486       C2H2-TYPE.
FT   ZN_FING     492    514       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     520    542       C2H2-TYPE.
SQ   SEQUENCE   574 AA;  65951 MW;  482E5F1176ECA5B0 CRC64;

  Query Match             17.9%;  Score 398;  DB 1;  Length 574;
  Best Local Similarity   30.2%;  Pred. No. 1.3e-16;
  Matches  121;  Conservative   48;  Mismatches  167;  Indels   64;  Gaps   18;

Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
              |  | | |  : | ||||  | |  | | |:  :|||  ::| ||| | ||| || : | 
Db        207 SGEKPYKC--KECGKAYNETSNLSTHKRIHTGKKPYKC--EECGKAFNRLSHLTTHKIIH 262

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
              : |||: |  |||  |   :|  |:  ||  | :||  | |  || :  :|  | |: 
Db        263 TGKKPYKCEECGKAFNQSANLTTHKRIHTGEKPYKC--EECGRAFSQSSTLTAHKIIHAG 320

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSH 194
```

```
Db        321 EKPYKCEECGKAFSQSSTLTTHKI-IHTGEKPYKCEE--CGKAFSQLSHLTTHKRIHSGE 377

Qy        195 PKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDG 254

Db        378 KPYKCEECGKAFKQSSTLTTHKRIHAGE---KFYKCEVCSKA-FSRFSHLTTHKRI-HTG 432

Qy        255 NIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDS 314

Db        433 EKP---YKCEECGKAFNLSSQLTTHKIIHTGE-KPYKCEEC----------GKAFNQSST 478

Qy        315 MSAQRSI-------------KSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYD 361

Db        479 LSKHKVIHTGEKPYKYEECGKAFNQS--SHLTTHKMIHTGEKPYKC--EECGKAFNNSSI 534

Qy        362 LRRHLKWHD-------------DNLQRIESFLNSIEKEE 387

Db        535 LNRHKMIHTGEKLYKPESCNNACDNIAKISKYKRNCAGEK 574
```

RESULT 5
TF3A_HUMAN
ID TF3A_HUMAN STANDARD; PRT; 423 AA.
AC Q92664; Q12963; Q13097;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Transcription factor IIIA (Factor A) (TFIIIA).
GN GTF3A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal brain;
RX MEDLINE=95309028; PubMed=7789179;
RA Arakawa H., Nagase H., Hayashi N., Ogawa M., Nagata M.,
RA Fujiwara T., Takahashi E., Shin S., Nakamura Y.;
RT "Molecular cloning, characterization, and chromosomal mapping of a
RT novel human gene (GTF3A) that is highly homologous to Xenopus
RT transcription factor IIIA.";
RL Cytogenet. Cell Genet. 70:235-238(1995).
RN [2]
RP SEQUENCE OF 61-423 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95347600; PubMed=7622052;
RA Drew P.D., Nagle J.W., Canning R.D., Ozato K., Biddison W.E.,
RA Becker K.G.;
RT "Cloning and expression analysis of a human cDNA homologous to
RT Xenopus TFIIIA.";
RL Gene 159:215-218(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=94342241; PubMed=8063702;
RA Moorefield B., Roeder R.G.;
RT "Purification and characterization of human transcription factor
RT IIIA.";
RL J. Biol. Chem. 269:20857-20865(1994).
CC -!- FUNCTION: INTERACTS WITH THE INTERNAL CONTROL REGION (ICR) OF
CC APPROXIMATELY 50 BASES WITHIN THE 5S RNA GENES, IS REQUIRED FOR
CC CORRECT TRANSCRIPTION OF THESE GENES BY RNA POLYMERASE III. ALSO
CC BINDS THE TRANSCRIBED 5S RNA'S. MAY INITIATE TRANSCRIPTION OF THE
CC 5S RIBOSOMAL RNA GENE AND MAINTAIN THE STABILITY OF TRANSCRIPTION
CC OF OTHER GENES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: Ubiquitous.
CC --
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC --
DR EMBL; D32257; BAA06988.1; -.
DR EMBL; U20272; AAA75623.1; -.
DR EMBL; U14134; AAA21873.1; -.
DR HSSP; P03001; 1TF3.
DR TRANSFAC; T04953; -.
DR Genew; HGNC:4662; GTF3A.
DR MIM; 600860; -.
DR GO; GO:0003709; F:RNA polymerase III transcription factor act. . .; TAS.
DR GO; GO:0009303; P:rRNA transcription; TAS.
DR GO; GO:0006383; P:transcription from Pol III promoter; TAS.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 9.
DR SMART; SM00355; ZnF_C2H2; 9.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
KW RNA-binding; Repeat; Nuclear protein; Polymorphism.

FT	ZN_FING	98	122 C2H2-TYPE.
FT	ZN_FING	128	152 C2H2-TYPE.
FT	ZN_FING	158	183 C2H2-TYPE.
FT	ZN_FING	190	212 C2H2-TYPE (ATYPICAL).
FT	ZN_FING	220	244 C2H2-TYPE.
FT	ZN_FING	247	271 C2H2-TYPE.
FT	ZN_FING	275	297 C2H2-TYPE.
FT	ZN_FING	304	329 C2H2-TYPE.
FT	ZN_FING	335	359 C2H2-TYPE.
FT	VARIANT	303	303 V -> L (in dbSNP:7323).
FT			/FTId=VAR_014824.
FT	CONFLICT	214	214 N -> H (IN REF. 2).
FT	CONFLICT	221	221 K -> KL (IN REF. 2; AAA21873).
FT	CONFLICT	223	248 MISSING (IN REF. 2; AAA21873).
FT	CONFLICT	378	387 SLASHLSGYI -> EFGLSSQWIY (IN REF. 1).

SQ SEQUENCE 423 AA; 46847 MW; A627D064A43FB6F0 CRC64;

Query Match 17.8%; Score 397.5; DB 1; Length 423;
Best Local Similarity 31.1%; Pred. No. 9.9e-17;
Matches 99; Conservative 55; Mismatches 95; Indels 69; Gaps 16;

```
Qy          8 KSISSL------ISSSSSSR--------PKKYICTYEGCDKAYNRPSLLEQHLRTHSNDR 53

Db         67 ESVSSLTIADAFIAAGESSAPTPPRPALPRRFICSFPDCSANYSKAWKLDAHLCKHTGER 126

Qy         54 PYKCTVDDCDKAFFRKSHLETHIVSHSEKKPFHCSV--CGKGVNSRQHLKRH-EITH--- 107

Db        127 PFVCDYEGCGKAFIRDYHLSRHILTHTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQ 186

Qy        108 TKSFKCTFENCQEAFYKHQSLR-HHILSVHEKTLTCKQ--CNKVFTRPSKLAQHKLKHHG 164

Db        187 QKQYICSFEDCKKTFKKHQQLKIHQCQNTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEG 246

Qy        165 GSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK---------------------------- 196

Db        247 ----YVC-QKGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPER 301

Qy        197 --LKCPK--CGKGCVGKKGLSSHMLS-HDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNI 250

Db        302 DVCRCPREGCGRTYTTVFNLQSHILSFHEES---RPFVCEHAGCGKTFAMKQSLTRH-AV 357

Qy        251 FHDGNIPDDLLKETEVKK 268

Db        358 VHD---PDKKKMKLKVKK 372
```

RESULT 6
ZN85_HUMAN
ID ZN85_HUMAN STANDARD; PRT; 595 AA.
AC Q03923;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```
DE   Zinc finger protein 85 (Zinc finger protein HPF4) (HTF1).
GN   ZNF85.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   TISSUE=Placenta;
RX   MEDLINE=99053537; PubMed=9839802;
RA   Poncelet D.A., Bellefroid E.J., Bastiaens P.V., Demoitie M.A.,
RA   Marine J.C., Pendeville H., Alami Y., Devos N., Lecocq P.J., Ogawa T.,
RA   Muller M., Martial J.A.;
RT   "Functional analysis of ZNF85 KRAB zinc finger protein, a member of
RT   the highly homologous ZNF91 family.";
RL   DNA Cell Biol. 17:931-943(1998).
RN   [2]
RP   SEQUENCE OF 1-196 FROM N.A.
RX   MEDLINE=91219421; PubMed=2023909;
RA   Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
RA   Martial J.A.;
RT   "The evolutionarily conserved Kruppel-associated box domain defines a
RT   subfamily of eukaryotic multifingered proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
CC   -!- FUNCTION: Transcriptional repressor.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Expressed preferentially in testicular
CC       tissues.
CC   -!- DEVELOPMENTAL STAGE: Expressed early during embryonic development.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 1 KRAB domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; U35376; AAA79179.1; -.
DR   EMBL; M61866; AAA52689.1; -.
DR   EMBL; M61868; AAA58671.1; -.
DR   PIR; C39384; C39384.
DR   PIR; G02075; G02075.
DR   HSSP; P08048; 7ZNF.
DR   TRANSFAC; T04990; -.
DR   Genew; HGNC:13160; ZNF85.
DR   MIM; 603899; -.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   GO; GO:0003714; F:transcription co-repressor activity; TAS.
DR   GO; GO:0003700; F:transcription factor activity; TAS.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 15.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 13.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 15.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 14.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 15.
KW   Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
KW   Nuclear protein; Repeat; Repressor.
FT   DOMAIN        4     75       KRAB.
FT   ZN_FING     146    168       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     174    196       C2H2-TYPE.
FT   ZN_FING     202    224       C2H2-TYPE.
FT   ZN_FING     230    252       C2H2-TYPE.
FT   ZN_FING     258    280       C2H2-TYPE.
FT   ZN_FING     286    308       C2H2-TYPE.
FT   ZN_FING     314    336       C2H2-TYPE.
FT   ZN_FING     342    364       C2H2-TYPE.
FT   ZN_FING     370    392       C2H2-TYPE.
FT   ZN_FING     398    420       C2H2-TYPE.
FT   ZN_FING     426    448       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     454    476       C2H2-TYPE.
FT   ZN_FING     482    504       C2H2-TYPE.
FT   ZN_FING     510    532       C2H2-TYPE.
FT   ZN_FING     538    560       C2H2-TYPE.
FT   ZN_FING     566    588       C2H2-TYPE.
FT   CONFLICT     84     84       R -> Q (IN REF. 2; AAA52689).
FT   CONFLICT    115    115       R -> I (IN REF. 2; AAA58671).
FT   CONFLICT    177    177       T -> R (IN REF. 2; AAA58671).
FT   CONFLICT    184    184       G -> R (IN REF. 2; AAA58671).
SQ   SEQUENCE   595 AA;  68718 MW;  44AA0A236D62D43B CRC64;

  Query Match             17.8%;  Score 396.5;  DB 1;  Length 595;
  Best Local Similarity   31.7%;  Pred. No. 1.6e-16;
  Matches  120;  Conservative   48;  Mismatches  160;  Indels   51;  Gaps   17;

Qy         10 ISSLISSSS-SSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFR 68
              ||  |   |  :|   | |  | | ||:|  | | :| | |: ::|||| ::| ||| :
Db        186 ISCLTEHSRIHTRVNFYKC--EECGKAFNWSSTLTKHKRIHTGEKPYKC--EECGKAFNQ 241

Qy         69 KSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTPENCQEAFYKHQ 126
               |:|  |   |: :||: |  ||| |     |  :| ||  | :|| : | :|| :
Db        242 SSNLIKHKKIHTGEKPYKCEECGKAFNRFSTLTTHKIIHTGEKPYKC--KECGKAFNRSS 299

Qy        127 SL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSV 185
             :|  |  :   ||   |::| | | : | |  ||: | | | |:|     | | |   :
Db        300 TLTTHRKIHTGEKPYKCEECGKAFKQSSNLTTHKIIHTGEKP-YKCKK--CGKAFNQSAH 356

Qy        186 LQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNE 243
              |  |  |        || ||||       |:| : |   | | : |  |  ||  | :
Db        357 LTTHEVIHTGEKPYKCEKCGKAFNHFSHLTTHKIIH---TGEKPYKCKEC--GKAFKHSS 411

Qy        244 LVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSL 303
              :  : | | |       |  : |: |   :| |||        ||  |:   |: |
Db        412 TLTKHKIIHTGE------KPYKSKECEKAFNQSSKL-------TEHKKIHTGEKPYE--- 455

Qy        304 DEKRSDVRSDSMSAQRSIKSFTASL--------EGSK-----SVSKLISNSGKKINCPKN 350
              ||     : | :   |  || |           :| |     :: |:|    |   |
Db        456 CEKCGKAFNQSSNLTRHKKSHTEEKPYKCEECGKGFKWPSTLTIHKIIHTGEKPYKC--E 513

Qy        351 NCDRMFSREYDLRRHLKWH 369
              | : |::   | :| | |
Db        514 ECGKAFNQSSKLTKHKKIH 532


RESULT 7
ZN93_HUMAN
ID   ZN93_HUMAN     STANDARD;      PRT;   616 AA.
AC   P35789; Q9Y2N8;
DT   01-JUN-1994 (Rel. 29, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein 93 (Zinc finger protein HTF34) (Fragment).
GN   ZNF93.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lamerdin J.E., McCready P.M., Skowronski E., Sakaldasis G.,
RA   Burkhart-Schultz K., Gordon L., Dias J., Scott D., Stilwagen S.,
RA   Phan H., Velasco N., Do L., Regala W., Terry A., Danganan L.,
RA   Erler A., Christensen M., Georgescu A., Avila J., Attix C.,
RA   Andreise T., Trankheim M., Amico-Keller G., Coefield J., Duarte S.,
RA   Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B., Arellano A.,
RA   Sanders C., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.S.,
RA   Carrano A.V.;
```

```
RT   "Sequence analysis of a 5.7 Mb region in 19p13.1 between OLFR and
RT   D19S455.";
RL   Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-191 FROM N.A.
RX   MEDLINE=91219421; PubMed=2023909;
RA   Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
RA   Martial J.A.;
RT   "The evolutionarily conserved Kruppel-associated box domain defines a
RT   subfamily of eukaryotic multifingered proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
CC   -!- FUNCTION: May be involved in transcriptional regulation.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- DEVELOPMENTAL STAGE: Expressed early during embryonic development.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 1 KRAB domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AC007204; AAD22981.1; -.
DR   EMBL; M61873; AAA83548.1; -.
DR   HSSP; P08046; 1A1H.
DR   Genew; HGNC:13169; ZNF93.
DR   MIM; 603975; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003700; F:transcription factor activity; NAS.
DR   GO; GO:0008270; F:zinc ion binding; NAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 16.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 7.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 16.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 16.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 17.
KW   Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
KW   Nuclear protein; Repeat.
FT   NON_TER       1      1
FT   DOMAIN       <1     71       KRAB.
FT   ZN_FING     141    163       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     169    191       C2H2-TYPE.
FT   ZN_FING     197    219       C2H2-TYPE.
FT   ZN_FING     225    247       C2H2-TYPE.
FT   ZN_FING     253    275       C2H2-TYPE.
FT   ZN_FING     281    303       C2H2-TYPE.
FT   ZN_FING     309    331       C2H2-TYPE.
FT   ZN_FING     337    359       C2H2-TYPE.
FT   ZN_FING     365    387       C2H2-TYPE.
FT   ZN_FING     393    415       C2H2-TYPE.
FT   ZN_FING     421    443       C2H2-TYPE.
FT   ZN_FING     449    471       C2H2-TYPE.
FT   ZN_FING     477    499       C2H2-TYPE.
FT   ZN_FING     505    527       C2H2-TYPE.
FT   ZN_FING     533    555       C2H2-TYPE.
FT   ZN_FING     561    583       C2H2-TYPE.
FT   ZN_FING     589    611       C2H2-TYPE.
FT   CONFLICT     71     71       S -> SGP (IN REF. 2).
FT   CONFLICT     85     85       Q -> H (IN REF. 2).
FT   CONFLICT    115    115       S -> R (IN REF. 2).
FT   CONFLICT    132    132       C -> S (IN REF. 2).
FT   CONFLICT    149    149       V -> D (IN REF. 2).
```

```
SQ   SEQUENCE   616 AA;  70572 MW;  DF309883AB61160A CRC64;

  Query Match             17.7%;  Score 395;  DB 1;  Length 616;
  Best Local Similarity   29.7%;  Pred. No. 2.1e-16;
  Matches  123;  Conservative   42;  Mismatches  157;  Indels   92;  Gaps   20;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
              | |||  | | ||:    | |  | | |: ::|||| | |||||   | |  | : |: |
Db        195 KPYIC--EECGKAFKYSSALNTHKRIHTGEKPYKC--DKCDKAFIASSTLSKHEIIHTGK 250

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHH--------- 131
              ||: |  |||  |    | :|:  ||  | :||  | | :|| :  :|  |
Db        251 KPYKCEECGKAFNQSSTLTKHKKIHTGEKPYKC--EECGKAFNQSSTLTKHKKIHTGEKP 308

Qy        132 -------------ILSVH------EKTLTCKQCNKVFTRPSKLAQHKLKH--------- 162
                          ||: |      ||   | :| | |   | |::|:  |
Db        309 YVCEECGKAFKYSRILTTHKRIHTGEKPYKCNKCGKAFIASSTLSRHEFIHMGKKHYKCE 368

Qy        163 -----------------HGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSH---PKLKCPK 201
                              | |   |:|:  | | |:  | |  | |:||      || :
Db        369 ECGKAFIWSSVLTRHKRVHTGEKPYKCEE--CGKAFKYSSTLSSH-KRSHTGEKPYKCEE 425

Qy        202 CGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGNIPDDL 260
              |||  |  || | :  |   |  | :|:  || | : : | :|  |  |  |
Db        426 CGKAFVASSTLSKHEIIH---TGKKPYKCEEC--GKAFNQSSSLTKHKKI-HTGEKP--- 476

Qy        261 LKETEVKKLENLLDQGSKLNNLHELET--EKLKVEEDEEDEEDS---LDEKRSDVRSDSM 315
              |  |  |  |  | | |    :: |  :  | ||  :    |   :  |:   |
Db        477 YKCEECGKAFN---QSSSLTKHKKIHTGEKPYKCEECGKAFNQSSTLIKHKKIHTREKPY 533

Qy        316 SAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
                :  |:|  |   :  |::   |  |   | : |:   | | | |
Db        534 KCEECGKAF--HLSTHLTTHKILHTGEKPYRC--RECGKAFNHSATLSSHKKIH 583
```

```
RESULT 8
ZN91_HUMAN
ID   ZN91_HUMAN     STANDARD;      PRT;  1191 AA.
AC   Q05481;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein 91 (Zinc finger protein HTF10) (HPF7).
GN   ZNF91.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93223677; PubMed=8467795;
RA   Bellefroid E.J., Marine J.C., Ried T., Lecocq P.J., Riviere M.,
RA   Amemiya C.T., Poncelet D.A., Coulie P.G., de Jong P.J.,
RA   Szpirer C., Ward D.C., Martial J.A.;
RT   "Clustered organization of homologous KRAB zinc-finger genes with
RT   enhanced expression in human T lymphoid cells.";
RL   EMBO J. 12:1363-1374(1993).
RN   [2]
RP   SEQUENCE OF 15-204 FROM N.A.
RX   MEDLINE=91219421; PubMed=2023909;
RA   Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
RA   Martial J.A.;
RT   "The evolutionarily conserved Kruppel-associated box domain defines a
RT   subfamily of eukaryotic multifingered proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 1 KRAB domain.
CC   -!- CAUTION: THE SEQUENCE FROM POSITION 1159 TO THE C-TERMINAL IS
CC       DERIVED FROM THE TRANSLATION OF AN ALU REPEAT.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; L11672; AAA59469.1; -.
DR   EMBL; M61871; AAA58672.1; ALT_SEQ.
DR   PIR; S35305; S35305.
DR   HSSP; P08047; 1SP2.
DR   Genew; HGNC:13166; ZNF91.
DR   MIM; 603971; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003700; F:transcription factor activity; NAS.
DR   GO; GO:0008270; F:zinc ion binding; NAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 34.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 20.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 33.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 31.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 35.
KW   Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
KW   Nuclear protein; Repeat.
FT   DOMAIN       13    84       KRAB.
FT   ZN_FING     154   176       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     179   200       C2H2-TYPE.
FT   ZN_FING     208   232       C2H2-TYPE.
FT   ZN_FING     238   260       C2H2-TYPE.
FT   ZN_FING     266   288       C2H2-TYPE.
FT   ZN_FING     294   316       C2H2-TYPE.
FT   ZN_FING     322   344       C2H2-TYPE.
FT   ZN_FING     350   372       C2H2-TYPE.
FT   ZN_FING     378   400       C2H2-TYPE.
FT   ZN_FING     406   428       C2H2-TYPE.
FT   ZN_FING     434   456       C2H2-TYPE.
FT   ZN_FING     462   484       C2H2-TYPE.
FT   ZN_FING     490   512       C2H2-TYPE.
FT   ZN_FING     518   540       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     546   568       C2H2-TYPE.
FT   ZN_FING     574   596       C2H2-TYPE.
FT   ZN_FING     602   624       C2H2-TYPE.
FT   ZN_FING     630   652       C2H2-TYPE.
FT   ZN_FING     658   680       C2H2-TYPE.
FT   ZN_FING     686   708       C2H2-TYPE.
FT   ZN_FING     714   736       C2H2-TYPE.
FT   ZN_FING     742   764       C2H2-TYPE.
FT   ZN_FING     770   792       C2H2-TYPE.
FT   ZN_FING     798   820       C2H2-TYPE.
FT   ZN_FING     826   848       C2H2-TYPE.
FT   ZN_FING     854   876       C2H2-TYPE.
FT   ZN_FING     885   904       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     910   932       C2H2-TYPE.
FT   ZN_FING     938   960       C2H2-TYPE.
FT   ZN_FING     966   988       C2H2-TYPE.
FT   ZN_FING     994  1016       C2H2-TYPE.
FT   ZN_FING    1022  1044       C2H2-TYPE.
FT   ZN_FING    1050  1072       C2H2-TYPE.
FT   ZN_FING    1078  1100       C2H2-TYPE.
FT   ZN_FING    1106  1128       C2H2-TYPE.
FT   ZN_FING    1134  1156       C2H2-TYPE.
SQ   SEQUENCE   1191 AA;  137225 MW;  581056BB11B8716D CRC64;

  Query Match           17.7%;  Score 393.5;  DB 1;  Length 1191;
  Best Local Similarity  35.8%;  Pred. No. 5.2e-16;
  Matches   98;  Conservative   40;  Mismatches  109;  Indels   27;  Gaps   12;

Qy        20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
             :| | | |  | | ||:::|| |  | | |: ::|||| ::| ||| : | | || : |
Db       905 TREKPYKC--EECGKAFSQPSHLTTHKRMHTGEKPYKC--EECGKAFSQSSTLTTHKIIH 960

Qy        80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
             : :||: |  |||       |  |:| ||  | :||  | | :|| :  :| ||  :
Db       961 TGEKPYKCEECGKAFRKSSTLTEHKIIHTGEKPYKC--EECGKAFSQSSTLTRHTRMHTG 1018

Qy       137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK 196
             ||   |::| | | | ||| ||: | |  | |:|:   | | | : | | | |: | :
Db      1019 EKPYKCEECGKAFNRSSKLTTHKIIHTGEKP-YKCEE--CGKAFISSSTLNGH-KRIHTR 1074

Qy       197 ---LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHD 253
               || :|||       |: |   |  | | : |  |  ||  |::  :  : | |
Db      1075 EKPYKCEECGKAFSQSSTLTRHKRLH---TGEKPYKCGEC--GKAFKESSALTKHKIIHT 1129

Qy       254 GNIPDDLLKETEVKKLENLLDQGSKLNNLHELET 287
             |      |  : :|     :| | | |  :: |
Db      1130 GE------KPYKCEKCCKAFNQSSILTNHKKIHT 1157


RESULT 9
Z347_HUMAN
ID   Z347_HUMAN     STANDARD;      PRT;   839 AA.
AC   Q96SE7; Q8TCN1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein 347 (Zinc finger 1111).
GN   ZNF347 OR ZNF1111.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Aitken C.J., Nicholson G.C.;
RT   "RANKL regulated zinc finger protein in osteoclastogenesis.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 305-839 FROM N.A.
RC   TISSUE=Brain;
RA   Bloecker H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: May function as a transcription factor.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 1 KRAB domain.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AY029765; AAK37403.1; -.
DR   EMBL; AL713691; CAD28491.1; -.
DR   Genew; HGNC:16447; ZNF347.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 18.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 2.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 19.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 17.
```

```
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 20.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat.
FT   DOMAIN        8     79       KRAB.
FT   ZN_FING     261    283       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     289    311       C2H2-TYPE.
FT   ZN_FING     317    339       C2H2-TYPE.
FT   ZN_FING     345    367       C2H2-TYPE.
FT   ZN_FING     373    395       C2H2-TYPE.
FT   ZN_FING     401    423       C2H2-TYPE.
FT   ZN_FING     429    451       C2H2-TYPE.
FT   ZN_FING     457    479       C2H2-TYPE.
FT   ZN_FING     485    507       C2H2-TYPE.
FT   ZN_FING     513    535       C2H2-TYPE.
FT   ZN_FING     541    563       C2H2-TYPE.
FT   ZN_FING     569    591       C2H2-TYPE.
FT   ZN_FING     597    619       C2H2-TYPE.
FT   ZN_FING     625    647       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     653    675       C2H2-TYPE.
FT   ZN_FING     681    703       C2H2-TYPE.
FT   ZN_FING     709    731       C2H2-TYPE.
FT   ZN_FING     737    759       C2H2-TYPE.
FT   ZN_FING     765    787       C2H2-TYPE.
SQ   SEQUENCE   839 AA;  95771 MW;  F7F7CC7069E7844E CRC64;

  Query Match             17.4%;  Score 387.5;  DB 1;  Length 839;
  Best Local Similarity   30.7%;  Pred. No. 8e-16;
  Matches  112;  Conservative   46;  Mismatches  154;  Indels   53;  Gaps   16;

Qy          11 SSL-ISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRK 69
               ||| |  :: | | | |     | | : : | |  | | |: ::|||| ::| ||| :
Db         386 SSLAIHQATHSGEKPYKC--NECGKVFTQNSHLTNHWRIHTGEKPYKC--NECGKAFGVR 441

Qy          70 SHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAPYKHQS 127
               | |  |:| |: :||: |  |||       || ||:: ||  | :||    | :|| | :
Db         442 SSLAIHLVIHTGEKPYKCHECGKVFRRNSHLARHQLIHTGEKPYKC--NECGKAFRAHSN 499

Qy         128 L-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVL 186
               |  | ::   ||   | :| |||: | || |:  | |  | | |:   | | |  :| |
Db         500 LTTHQVIHTGEKPYKCNECGKVFTQNSHLANHQRIHTGVKP-YMCNE--CGKAFSVYSSL 556

Qy         187 QFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNEL 244
                  |  |        || :|||       |: |   |   | | : |: |  ||  : |
Db         557 TTHQVIHTGEKPYKCNECGKVFTQNSHLARHRGIH---TGEKPYKCNEC--GKVFRHNSY 611

Qy         245 VEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLD 304
               :  :   | | |   | | |     : | | |     : | :   : :|
Db         612 LSRHQRIHTGEKP---YKYNEYGK---AFSEHSNLTTHQVIHTGEKPYKCNE-------- 657

Qy         305 EKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSREYDLRR 364
                         | || :   ::   : :   ||   |  | | : ||:   | |
Db         658 --------------CGKVFTQNSHLAR--HRRVHTGGKPYQC--NECGKAFSQTSKLAR 698

Qy         365 HLKWH 369
               | : |
Db         699 HQRVH 703


RESULT 10
Z431_HUMAN
ID   Z431_HUMAN     STANDARD;      PRT;   576 AA.
AC   Q8TF32; Q8IWC4;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein 431.
GN   ZNF431 OR KIAA1969.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=21842142; PubMed=11853319;
RA   Nagase T., Kikuno R., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XXII.
RT   The complete sequences of 50 new cDNA clones which code for large
RT   proteins.";
RL   DNA Res. 8:319-327(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length human
RT   and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: May be involved in transcriptional regulation.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 1 KRAB domain.
CC   -!- SIMILARITY: Contains 13 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AB075849; BAB85555.1; ALT_INIT.
DR   EMBL; BC040506; AAH40506.1; -.
DR   Genew; HGNC:20809; ZNF431.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 12.
DR   PRINTS; PR00048; ZINCFINGER.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 12.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 12.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 13.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat.
FT   DOMAIN       35    106       KRAB.
FT   ZN_FING     176    198       C2H2-TYPE 1 (DEGENERATE).
FT   ZN_FING     204    226       C2H2-TYPE 2.
FT   ZN_FING     232    254       C2H2-TYPE 3.
FT   ZN_FING     260    282       C2H2-TYPE 4.
FT   ZN_FING     288    310       C2H2-TYPE 5.
FT   ZN_FING     316    338       C2H2-TYPE 6.
FT   ZN_FING     344    366       C2H2-TYPE 7.
FT   ZN_FING     372    394       C2H2-TYPE 8.
FT   ZN_FING     400    422       C2H2-TYPE 9.
FT   ZN_FING     428    450       C2H2-TYPE 10.
```

```
FT   ZN_FING     456    478       C2H2-TYPE 11.
FT   ZN_FING     484    506       C2H2-TYPE 12.
FT   ZN_FING     512    534       C2H2-TYPE 13.
FT   CONFLICT    198    198       H -> R (IN REF. 2).
SQ   SEQUENCE   576 AA;  67216 MW;  532774BF69EC9E2A CRC64;

  Query Match             17.3%;  Score 385;  DB 1;  Length 576;
  Best Local Similarity   29.4%;  Pred. No. 7.5e-16;
  Matches  109;  Conservative   48;  Mismatches  158;  Indels   56;  Gaps   15;

Qy          8 KSISSLISSSSSS----RPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCD 63
              ||    |: |         |    | |  | | ||:   | | :| | |: ::|:||  ::|
Db        211 KSFCMLLHLSQHKRIHIRENSYQC--EECGKAFKWFSTLTRHKRIHTGEKPFKC--EECG 266

Qy         64 KAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEA 121
              ||| : | | || : |: :||: |   |||   |    ||   |:| ||   | :||   | | :|
Db        267 KAFKQSSTLTTHKIIHTGEKPYRCEECGKAFNRSSHLTTHKIIHTGEKPYKC--EECGKA 324

Qy        122 FYKHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNF 180
              | :  :|  |  :   ||   |::|:| | | | | :||:  | |  :|:|:   | | |
Db        325 FNQSSTLSTHKFIHAGEKPYKCEECDKAFNRFSYLTKHKI-IHTGEKSYKCEE--CGKGF 381

Qy        181 QTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKF 238
              | |  |  |         ||  |||       |::| :  |   |  | : |: |  ||
Db        382 NWSSTLTKHKRIHTGEKPYKCEVCGKAFNESSNLTTHKMIH---TGEKPYKCEEC--GKA 436

Qy        239 AKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEED 298
              ::  :  : | | |       |  : ::      | | |
Db        437 FNRSPQLTAHKIIHTGE------KPYKCEECGKAFSQSSILTT---------------- 473

Qy        299 EEDSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDRMFSR 358
                          ||        :  |:|  |   :|   |:|   |   |   | : |::
Db        474 ------HKRIHTGEKPYKCEECGKAFNRSSNLTK--HKIIHTGEKSYKC--EECGKAFNQ 523

Qy        359 EYDLRRHLKWH 369
                | :| | |
Db        524 SSTLTKHRKIH 534


RESULT 11
Z228_HUMAN
ID   Z228_HUMAN     STANDARD;      PRT;   913 AA.
AC   Q9UJU3; Q9HCA7;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein 228.
GN   ZNF228.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Shannon M., Branscomb E., Hauser L., Gordon L., Ashworth L.,
RA   Stubbs L.;
RT   "Differential expansion of homologous zinc-finger gene families in
RT   human chromosome 19q13.2 and mouse chromosome 7.";
RL   Submitted (OCT-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Kodoyianni V., Ge Y., Krummel G.K., Kvikstad E., Grable L.,
RA   Severin J., Gordon L., Shannon M., Brower A., Olsen A.S., Smith L.M.;
RT   "Sequence analysis of a 1Mb region in 19q13.2 containing a zinc finger
RT   gene cluster.";
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: May function as a transcription factor.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 1 KRAB domain.
CC   -!- SIMILARITY: Contains 17 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF198358; AAF12816.1; -.
DR   EMBL; AC084239; AAG23968.1; -.
DR   HSSP; P08047; 1SP2.
DR   Genew; HGNC:13021; ZNF228.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 13.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 12.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 13.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 13.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 17.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein.
FT   DOMAIN        8     79       KRAB.
FT   ZN_FING     258    280       C2H2-TYPE 1 (DEGENERATE).
FT   ZN_FING     443    465       C2H2-TYPE 2 (DEGENERATE).
FT   ZN_FING     471    493       C2H2-TYPE 3 (DEGENERATE).
FT   ZN_FING     497    519       C2H2-TYPE 4 (DEGENERATE).
FT   ZN_FING     525    547       C2H2-TYPE 5.
FT   ZN_FING     553    575       C2H2-TYPE 6.
FT   ZN_FING     581    603       C2H2-TYPE 7.
FT   ZN_FING     609    631       C2H2-TYPE 8.
FT   ZN_FING     637    659       C2H2-TYPE 9.
FT   ZN_FING     665    687       C2H2-TYPE 10.
FT   ZN_FING     693    715       C2H2-TYPE 11.
FT   ZN_FING     721    743       C2H2-TYPE 12.
FT   ZN_FING     749    771       C2H2-TYPE 13.
FT   ZN_FING     777    799       C2H2-TYPE 14.
FT   ZN_FING     805    827       C2H2-TYPE 15.
FT   ZN_FING     833    855       C2H2-TYPE 16.
FT   ZN_FING     861    883       C2H2-TYPE 17.
FT   CONFLICT      3      5       VSK -> KFQ (IN REF. 2).
FT   CONFLICT    113    113       T -> A (IN REF. 2).
FT   CONFLICT    143    143       A -> V (IN REF. 2).
FT   CONFLICT    222    222       E -> K (IN REF. 2).
FT   CONFLICT    261    261       S -> T (IN REF. 2).
FT   CONFLICT    392    392       S -> N (IN REF. 2).
FT   CONFLICT    446    446       Q -> E (IN REF. 2).
FT   CONFLICT    485    485       H -> Y (IN REF. 2).
FT   CONFLICT    813    813       A -> G (IN REF. 2).
SQ   SEQUENCE   913 AA;  105702 MW;  8D19F5D27CD9FA6B CRC64;

  Query Match             17.2%;  Score 383;  DB 1;  Length 913;
  Best Local Similarity   30.7%;  Pred. No. 1.6e-15;
  Matches  109;  Conservative   55;  Mismatches  133;  Indels   58;  Gaps   19;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
              | | |  | ||| ::| | |: | | | |: ::||||  ::| | | | |:|: |    |: :
Db        551 KPYKC--EECDKGFSRSSYLQAHQRVHTGEKPYKC--EECGKGFSRNSYLQGHQRVHTGE 606

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLR-HHILSVHEKT 139
              ||: |  |||| :   ||: |:  ||  | ||| | | : |    :|: |  :   ||
Db        607 KPYKCEECGKGFSRSSHLQGHQRVHTGEKPFKC--EECGKGFSWSFNLQIHQRVHTGEKP 664

Qy        140 LTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPKL 197
                |::| | |:: | | |:  | | | |||   | |:|   | ||| |  :
Db        665 YKCEECGKGFSKASTLLAHQRVHTGEKP-YQCDE--CGKSFSQRSYLQSHQSVHSGERPY 721

Qy        198 KCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDGNIP 257
              |  ||||   :  |  |   |  | :| : |: |  | |:: : |  |  : | |  |
```

```
Db        722 ICEVCGKGFSQRAYLQGHQRVH---TRVKPYKCEMCGKG-FSQSSRLEAHRRV-HTGGKP 776

Qy        258 DDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDSMSA 317

Db        777 ---------YKCE-VCTKGFS-------ESSRLQAHQ------------RVHVEGRPYKC 807

Qy        318 QRSIKSFT--ASLEGSKSVSKLISNSGKK-INCPKNNCDRMFSREYDLRRHLKWH 369

Db        808 EQCGKAFSGYSSLQAHHRV-----HTGEKPYKC--EVCGKGFSQRSNLQAHQRVH 855


RESULT 12
Z430_HUMAN
ID   Z430_HUMAN     STANDARD;      PRT;   570 AA.
AC   Q9H8G1; Q86V70;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein 430.
GN   ZNF430.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Bi A., Yu L.;
RT   "Homo sapiens mRNA similar to zinc finger protein 85.";
RL   Submitted (APR-2003) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 1-364 FROM N.A.
RC   TISSUE=Placenta;
RA   Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA   Nishikawa T., Nagai K., Sato H., Sugano S., Shiratori A., Sudo H.,
RA   Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA   Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA   Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA   Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA   Ninomiya K., Iwayanagi T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: May be involved in transcriptional regulation.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 1 KRAB domain.
CC   -!- SIMILARITY: Contains 12 C2H2-type zinc fingers.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AY269787; AAP30885.1; -.
DR   EMBL; AK023721; BAB14656.1; -.
DR   Genew; HGNC:20808; ZNF430.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 3.
DR   ProDom; PD000003; Znf_C2H2; 3.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 3.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 11.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat.
FT   DOMAIN       35    107       KRAB.
FT   ZN_FING     205    227       C2H2-TYPE 1.
FT   ZN_FING     233    255       C2H2-TYPE 2.
FT   ZN_FING     261    283       C2H2-TYPE 3.
FT   ZN_FING     289    311       C2H2-TYPE 4.
FT   ZN_FING     317    339       C2H2-TYPE 5.
FT   ZN_FING     345    367       C2H2-TYPE 6.
FT   ZN_FING     373    395       C2H2-TYPE 7.
FT   ZN_FING     401    423       C2H2-TYPE 8.
FT   ZN_FING     429    451       C2H2-TYPE 9.
FT   ZN_FING     457    479       C2H2-TYPE 10.
FT   ZN_FING     485    507       C2H2-TYPE 11.
FT   ZN_FING     513    535       C2H2-TYPE 12 (DEGENERATE).
FT   CONFLICT    293    364       ECGKTFNRSSHLTTHKRIHTGEKPYRCEECGRAFNRSSHLT
FT                                THKIIHTGEKPYKCEECGKAFNQSSTLTTHK -> NVAKPL
FT                                TGPHTLLHIKEFILERNPTDVKNVAELLTGPHTLLHIRLFI
FT                                LERNPTNVKNVAKLLTNPQPLLHIR (IN REF. 2).
SQ   SEQUENCE   570 AA;  66346 MW;  0561E871DC92B61F CRC64;

  Query Match             16.9%;  Score 376;  DB 1;  Length 570;
  Best Local Similarity   32.5%;  Pred. No. 2.5e-15;
  Matches   94;  Conservative   38;  Mismatches  109;  Indels   48;  Gaps   13;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82

Db        287 KPYRC--EECGKTFNRSSHLTTHKRIHTGEKPYRC--EECGRAFNRSSHLTTHKIIHTGE 342

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVHEK- 138

Db        343 KPYKCEECGKAFNQSSTLTTHKIIHAGEKPYKC--EECGKAFYRFSYLTKHKIIHTGEKF 400

Qy        139 -------------TLT-------------CKQCNKVFTRPSKLAQHKLKHHGGSPAYQC 171

Db        401 YKCEECGKGFNWSSTLTKHKRIHTGEKPYKCEQCGKAFNESSNLTAHKIIHTGEKP-YKC 459

Qy        172 DHPGCFKNFQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWT 229

Db        460 EE--CGKAFNRSPKLTAHKVIHSGEKPYKCEECGKAFNQFSNLTKHKITHIGDTSYKYLE 517

Qy        230 CDYCDVGKFAKKNELVEHYNIFHDGNIPDDLLKETEV-KKLENLLDQGS 277

Db        518 CDKA----FSQSSTLTKH-KVIHTGEKPYNCEEYGKAFNQSSNLIEQSN 561


RESULT 13
Z184_HUMAN
ID   Z184_HUMAN     STANDARD;      PRT;   751 AA.
AC   Q99676; O60792;
DT   15-JUL-1998 (Rel. 36, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein 184.
GN   ZNF184.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Phillips S.;
RL   Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 26-751 FROM N.A.
RC   TISSUE=Placenta;
RX   MEDLINE=97230463; PubMed=9073517;
RA   Goldwurm S., Menzies M.L., Banyer J.L., Powell B.L.W.,
RA   Jazwinska E.C.;
RT   "Identification of a novel Krueppel-related zinc finger gene (ZNF184)
RT   mapping to 6p21.3.";
RL   Genomics 40:486-489(1997).
CC   -!- FUNCTION: May be involved in transcriptional regulation.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- TISSUE SPECIFICITY: PREDOMINANT EXPRESSION IN TESTIS.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
```

```
CC   -!- SIMILARITY: Contains 1 KRAB domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AL021918; CAA17278.1; -.
DR   EMBL; U66561; AAC51180.1; -.
DR   HSSP; P08046; 1A1H.
DR   Genew; HGNC:12975; ZNF184.
DR   MIM; 602277; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003677; F:DNA binding; NAS.
DR   GO; GO:0008270; F:zinc ion binding; NAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 19.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 18.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 19.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 19.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 19.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat.
FT   DOMAIN       28     99       KRAB.
FT   ZN_FING     222    244       C2H2-TYPE.
FT   ZN_FING     250    272       C2H2-TYPE.
FT   ZN_FING     278    300       C2H2-TYPE.
FT   ZN_FING     306    328       C2H2-TYPE.
FT   ZN_FING     334    356       C2H2-TYPE.
FT   ZN_FING     362    384       C2H2-TYPE.
FT   ZN_FING     390    412       C2H2-TYPE.
FT   ZN_FING     418    440       C2H2-TYPE.
FT   ZN_FING     446    468       C2H2-TYPE.
FT   ZN_FING     474    496       C2H2-TYPE.
FT   ZN_FING     502    524       C2H2-TYPE.
FT   ZN_FING     530    552       C2H2-TYPE.
FT   ZN_FING     558    580       C2H2-TYPE.
FT   ZN_FING     586    608       C2H2-TYPE.
FT   ZN_FING     614    636       C2H2-TYPE.
FT   ZN_FING     642    664       C2H2-TYPE.
FT   ZN_FING     670    692       C2H2-TYPE.
FT   ZN_FING     698    720       C2H2-TYPE.
FT   ZN_FING     726    748       C2H2-TYPE.
FT   CONFLICT     27     27       A -> S (IN REF. 2).
FT   CONFLICT    295    295       Q -> R (IN REF. 2).
FT   CONFLICT    359    359       E -> G (IN REF. 2).
FT   CONFLICT    379    381       QHQ -> PHP (IN REF. 2).
FT   CONFLICT    498    498       R -> G (IN REF. 2).
FT   CONFLICT    526    526       Q -> G (IN REF. 2).
FT   CONFLICT    533    533       K -> E (IN REF. 2).
FT   CONFLICT    547    549       KHE -> QHQ (IN REF. 2).
FT   CONFLICT    746    746       R -> K (IN REF. 2).
SQ   SEQUENCE   751 AA;  86174 MW;  249DFB302C6BC97D CRC64;

  Query Match             16.9%;  Score 376;  DB 1;  Length 751;
  Best Local Similarity   25.2%;  Pred. No. 3.4e-15;
  Matches  114;  Conservative   52;  Mismatches  146;  Indels  140;  Gaps   20;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEK 82
              | | |  : | |  :     | || | |: ::|||  |:| ||| :::||  |    |: :
Db        276 KPYKC--DQCGKGFIEGPSLTQHQRIHTGEKPYKC--DECGKAFSQRTHLVQHQRIHTGE 331

Qy         83 KPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYK------HQS------- 127
```

```
              ||: |: |||  : | |   |:  ||  | |||  : | : | :        ||
Db        332 KPYTCNECGKAFSQRGHFMEHQKIHTGEKPFKC--DECDKTFTRSTHLTQHQKIHTGEKT 389

Qy        128 ----------------LRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQC 171
                           :|||::   ||   | :| | |:: | | ||:  | | | | | |
Db        390 YKCNECGKAFNGPSTFIRHHMIHTGEKPYECNECGKAFSQHSNLTQHQKTHTGEKP-YDC 448

Qy        172 DHPGCFKNFQTWSVLQFHIK----------------------QSHPKL-------KCPK 201
                | |:|  || |  |:|                        | ::       :| :
Db        449 AE--CGKSFSYWSSLAQHLKIHTGEKPYKCNECGKAFSYCSSLTQHRRIHTREKPFECSE 506

Qy        202 CGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGNIP--- 257
              |||         |: |  :|   |  | : |  |  || | : : | :|  | | |  |
Db        507 CGKAFSYLSNLNQHQKTH---TQEKAYECKEC--GKAFIRSSSLAKHERI-HTGEKPYQC 560

Qy        258 ----------DDLLKETEVKKLE---NLLDQGSKLN-NLHELETEKL------------- 290
                       |::  ::   |      : |   | |:|  : :::
Db        561 HECGKTFSYGSSLIQHRKIHTGERPYKCNECGRAFNQNIHLTQHKRIHTGAKPYECAECG 620

Qy        291 ------------KVEEDEEDEEDSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKL 337
                         |   :|:  : :  ||      |     ||
Db        621 KAFRHCSSLAQHQKTHTEEKPYQCNKCEKTFSQSSHLTQHQR------------------ 662

Qy        338 ISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
              |    |   | | ||: |||   |  |   |
Db        663 IHTGEKPYKC--NECDKAFSRSTHLTEHQNTH 692
```

```
RESULT 14
XFIN_XENLA
ID   XFIN_XENLA     STANDARD;      PRT;  1350 AA.
AC   P08045;
DT   01-AUG-1988 (Rel. 08, Created)
DT   01-AUG-1988 (Rel. 08, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein Xfin.
GN   XFIN.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Embryo;
RX   MEDLINE=88082679; PubMed=2826129;
RA   Ruiz i Altaba A., Perry-O'Keefe H., Melton D.A.;
RT   "Xfin: an embryonic gene encoding a multifingered protein in
RT   Xenopus.";
RL   EMBO J. 6:3065-3070(1987).
RN   [2]
RP   CHARACTERIZATION.
RX   MEDLINE=94021366; PubMed=7692399;
RA   Andreazzoli M., de Lucchini S., Costa M., Barsacchi G.;
RT   "RNA binding properties and evolutionary conservation of the Xenopus
RT   multifinger protein Xfin.";
RL   Nucleic Acids Res. 21:4218-4225(1993).
RN   [3]
RP   STRUCTURE BY NMR OF FINGER 31.
RX   MEDLINE=89346749; PubMed=2503871;
RA   Lee M.S., Gippert G.P., Soman K.V., Case D.A., Wright P.E.;
RT   "Three-dimensional solution structure of a single zinc finger DNA-
RT   binding domain.";
RL   Science 245:635-637(1989).
RN   [4]
RP   STRUCTURE BY NMR OF A FINGER.
RX   MEDLINE=89378224; PubMed=2506074;
RA   Lee M.S., Cavanagh J., Wright P.E.;
RT   "Complete assignment of the 1H NMR spectrum of a synthetic zinc
RT   finger from Xfin. Sequential resonance assignments and secondary
RT   structure.";
RL   FEBS Lett. 254:159-164(1989).
CC   -!- FUNCTION: BINDS RNA. COULD FUNCTION IN POST-TRANSLATIONAL
```

```
CC       REGULATION PROCESSES.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: OOCYTES AND IN SPECIALIZED CELL TYPES SUCH
CC       AS NEURAL RETINA CONES.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT OOGENESIS AND
CC       EMBRYOGENESIS.
CC   -!- DOMAIN: CONTAINS 37 FINGER MOTIFS IN 6 DOMAINS.
CC   -!- PTM: Phosphorylated.
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 1 KRAB domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X06021; CAA29425.1; -.
DR   PDB; 1ZNF; 15-OCT-91.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 36.
DR   ProDom; PD000003; Znf_C2H2; 20.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 35.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 35.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 37.
KW   Zinc-finger; Metal-binding; RNA-binding; Repeat; 3D-structure;
KW   Phosphorylation.
FT   DOMAIN       1     58       KRAB.
FT   ZN_FING    108    130       C2H2-TYPE.
FT   ZN_FING    136    158       C2H2-TYPE.
FT   ZN_FING    164    186       C2H2-TYPE.
FT   ZN_FING    192    214       C2H2-TYPE.
FT   ZN_FING    220    242       C2H2-TYPE.
FT   ZN_FING    248    270       C2H2-TYPE.
FT   ZN_FING    276    298       C2H2-TYPE.
FT   ZN_FING    326    348       C2H2-TYPE.
FT   ZN_FING    354    376       C2H2-TYPE.
FT   ZN_FING    382    404       C2H2-TYPE.
FT   ZN_FING    410    432       C2H2-TYPE.
FT   ZN_FING    438    460       C2H2-TYPE.
FT   ZN_FING    466    488       C2H2-TYPE.
FT   ZN_FING    503    525       C2H2-TYPE.
FT   ZN_FING    531    553       C2H2-TYPE.
FT   ZN_FING    559    581       C2H2-TYPE.
FT   ZN_FING    587    609       C2H2-TYPE.
FT   ZN_FING    615    637       C2H2-TYPE.
FT   ZN_FING    643    665       C2H2-TYPE.
FT   ZN_FING    671    693       C2H2-TYPE.
FT   ZN_FING    699    721       C2H2-TYPE.
FT   ZN_FING    750    772       C2H2-TYPE.
FT   ZN_FING    778    800       C2H2-TYPE.
FT   ZN_FING    806    828       C2H2-TYPE.
FT   ZN_FING    834    856       C2H2-TYPE.
FT   ZN_FING    862    884       C2H2-TYPE.
FT   ZN_FING    890    912       C2H2-TYPE.
FT   ZN_FING    918    940       C2H2-TYPE.
FT   ZN_FING    988   1010       C2H2-TYPE.
FT   ZN_FING   1016   1038       C2H2-TYPE.
FT   ZN_FING   1044   1066       C2H2-TYPE.
FT   ZN_FING   1136   1158       C2H2-TYPE.
FT   ZN_FING   1164   1186       C2H2-TYPE.
FT   ZN_FING   1192   1214       C2H2-TYPE.
FT   ZN_FING   1220   1242       C2H2-TYPE.
FT   ZN_FING   1248   1270       C2H2-TYPE.
FT   ZN_FING   1276   1298       C2H2-TYPE.
FT   STRAND    1045   1045
FT   STRAND    1052   1052
FT   HELIX     1056   1062
FT   HELIX     1063   1065
FT   TURN      1066   1066
SQ   SEQUENCE   1350 AA;  155804 MW;  27F10AB0851E0AD8 CRC64;

  Query Match             16.8%;  Score 375.5;  DB 1;  Length 1350;
  Best Local Similarity   27.3%;  Pred. No. 6.8e-15;
  Matches  119;  Conservative   65;  Mismatches  171;  Indels   81;  Gaps   20;

Qy         11 SSLISSSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKS 70
              | |    : || | |:|::  |:| : : | | :| |||: :|||:|   :| | | :||
Db        340 SFLNHQQTHSREKPYLCSH--CNKGFIQNSDLVKHFRTHTGERPYQCA--ECHKGFIQKS 395

Qy         71 HLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL 128
               |  |: :|: :||| || | |    | | :|: |||  | :||:  :| : | :  :|
Db        396 DLVKHLRTHTGEKPFKCSHCDKKFTERSALAKHQRTHTGEKPYKCS--DCGKEFTQRSNL 453

Qy        129 -RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKH--------HGGSPAYQCDHPGCFKN 179
               |  :   |:  |  |::| : | | :|: |          |  :  ::|   |
Db        454 ILHQRIHTGERPYKCTLCDRTFIQNSDLVKHQKVHANLPLSDPHTANSPHKCSK--CDLT 511

Qy        180 FQTWSVLQFHIK--QSHPKLKCPKCGKGCVGKKGLSSHMLSH----------------DD 221
              | ||   | |       | :| :| ||   |  |  |:  |                  :
Db        512 FSHWSTFMKHSKLHSGEKKFQCAECKKGFTQKSDLVKHIRVHTGEKPFKCLLCKKSFSQN 571

Qy        222 STMIKIW---------TCDYCDVGKFAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENL 272
              | : | |         |  ||   | ::: |::|:  | | |       :   :
Db        572 SDLHKHWRIHTGEKPFPCYTCD-KSFTERSALIKHHRT-HTGERPH------KCSVCQKG 623

Qy        273 LDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDSMSAQR-----------S 320
              | | |         |:  :    |:    :    |      || :  ||
Db        624 FIQKSAL-------TKHSRTHTGEKPYPCTQCGKSFIQNSDLVKHQRIHTGEKPYHCTEC 676

Qy        321 IKSFTASLEGSKSVSKLISNSGKK-INCPKNNCDRMFSREYDLRRHLKWHD-DNLQRIES 378
              | ||   |||  |    ::||:|   ||:  |:: | :  || :||  |: :|    :
Db        677 NKRFT---EGSSLVKHRRTHSGEKPYRCPQ--CEKTFIQSSDLVKHLVVHNGENPPAATA 731

Qy        379 FLN-SIEKEETPEGEP 393
              |   | :|    ||
Db        732 FHEILIRRENLTRSEP 747


RESULT 15
Z257_HUMAN
ID   Z257_HUMAN     STANDARD;      PRT;   535 AA.
AC   Q9Y2Q1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Zinc finger protein 257 (Bone marrow zinc finger 4) (BMZF-4).
GN   ZNF257 OR BMZF4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone marrow;
RX   MEDLINE=20054457; PubMed=10585455;
RA   Han Z.-G., Zhang Q.-H., Ye M., Kan L.-X., Gu B.-W., He K.-L.,
RA   Shi S.-L., Zhou J., Fu G., Mao M., Chen S.-J., Yu L., Chen Z.;
RT   "Molecular cloning of six novel Kruppel-like zinc finger genes from
RT   hematopoietic cells and identification of a novel transregulatory
RT   domain KRNB.";
RL   J. Biol. Chem. 274:35741-35748(1999).
CC   -!- FUNCTION: May function as a transcription factor.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC   -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC       FINGER PROTEINS.
CC   -!- SIMILARITY: Contains 1 KRAB domain.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AF070651; AAD20957.1; -.
DR   HSSP; P08047; 1SP2.
DR   Genew; HGNC:13498; ZNF257.
DR   MIM; 606957; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003677; F:DNA binding; NAS.
DR   GO; GO:0008270; F:zinc ion binding; NAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   InterPro; IPR001909; KRAB.
DR   InterPro; IPR007087; Znf_C2H2.
DR   InterPro; IPR007086; Znf_C2H2_sub.
DR   Pfam; PF01352; KRAB; 1.
DR   Pfam; PF00096; zf-C2H2; 10.
DR   PRINTS; PR00048; ZINCFINGER.
DR   ProDom; PD000003; Znf_C2H2; 9.
DR   SMART; SM00349; KRAB; 1.
DR   SMART; SM00355; ZnF_C2H2; 12.
DR   PROSITE; PS50805; KRAB; 1.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 10.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 12.
KW   Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
KW   Nuclear protein; Repeat.
FT   DOMAIN        4     75       KRAB.
FT   ZN_FING     173    195       C2H2-TYPE.
FT   ZN_FING     201    223       C2H2-TYPE.
FT   ZN_FING     229    251       C2H2-TYPE.
FT   ZN_FING     257    282       C2H2-TYPE (DEGENERATE).
FT   ZN_FING     288    310       C2H2-TYPE.
FT   ZN_FING     316    338       C2H2-TYPE.
FT   ZN_FING     344    366       C2H2-TYPE.
FT   ZN_FING     372    397       C2H2-TYPE (ATYPICAL).
FT   ZN_FING     403    425       C2H2-TYPE.
FT   ZN_FING     431    453       C2H2-TYPE.
FT   ZN_FING     459    481       C2H2-TYPE.
FT   ZN_FING     487    509       C2H2-TYPE.
SQ   SEQUENCE   535 AA;  62348 MW;  22DC5B0C4613BC51 CRC64;

  Query Match             16.8%;  Score 374;  DB 1;  Length 535;
  Best Local Similarity   33.5%;  Pred. No. 3.1e-15;
  Matches   88;  Conservative   38;  Mismatches  115;  Indels   22;  Gaps    9;

Qy          8 KSISSLISSSSSSRPK-----KYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDC 62
              |:       |:| | |     |  |  : | ||:|: | | :|  ||: ::||||  ::|
Db        149 KYVKVPYKFSNSDRHKIRHTEKKTCKCKECGKAFNQSSALTRHKMTHTGEKPYKC--EEC 206

Qy         63 DKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFENCQEAF 122
              ||| | |||  | | |: :||: |  |||  |   |: :|:  | :     :: | :||
Db        207 GKAFNRSSHLTQHKVIHTREKPYKCEECGKAFNRSSHITQHKRIHNREKPFKYDECCKAF 266

Qy        123 YKHQSL----RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFK 178
                :|    :|  :   ||   |::| | | : | | :||: | |  | :||:   | |
Db        267 KWSSALTTLTQHKRIHTGEKPYKCEECGKAFNQSSALTRHKMIHTGEKP-FQCEE--CGK 323

Qy        179 NFQTWSVLQFHIKQSHPK---LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCD-YCD 234
              |  | | | | | | |    || :|||      |: |   |   | | : || ||
Db        324 AFNRSSHLTQH-KIIHTKEKPYKCEECGKAFNRSSHLTKHKRIH---TREKAYKCDEYCK 379

Qy        235 VGKFAKKNELVEHYNIFHDGNIP 257
              ::     :  : | | |  |
Db        380 AFNWSSALTTLTQHKIIHTGEKP 402


Search completed: May  5, 2004, 15:21:55
Job time : 19 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 5, 2004, 15:19:16 ; Search time 20 Seconds
(without alignments)
1981.545 Million cell updates/sec

Title: US-09-831-804-3
Perfect score: 2229
Sequence: 1 MSESDETKSISSLISSSSSS.....PLVKKARMDLLPNETSVISR 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	837.5	37.6	429	2	S20050	transcription fact
2	495	22.2	374	2	T37676	zinc-finger protei
3	400	17.9	803	2	S26823	zinc finger protei
4	396.5	17.8	595	2	G02075	transcription repr
5	395.5	17.7	363	2	I38937	DNA/RNA-binding pr
6	393.5	17.7	1191	2	S35305	zinc finger protei
7	375.5	16.8	1350	2	S00647	zinc finger protein - A
8	374.5	16.8	710	2	I48668	zinc finger protei
9	373	16.7	542	2	A54661	zinc finger protei
10	373	16.7	651	2	B32891	finger protein 2,
11	372	16.7	393	2	JN0533	finger protein pML
12	364.5	16.4	338	2	G01496	transcription fact
13	364	16.3	428	2	A32891	finger protein 1,
14	363.5	16.3	615	2	S06546	finger protein (cl
15	359.5	16.1	427	2	A35659	krueppel-related p
16	354	15.9	693	2	I37570	zinc finger protei
17	353.5	15.9	591	2	S65088	finger protein XFO
18	351	15.7	474	2	I54338	zinc finger protei
19	348.5	15.6	439	2	S06556	finger protein (cl
20	348.5	15.6	707	2	S68858	finger protein - m
21	345.5	15.5	335	2	JC1441	transcription fact
22	345.5	15.5	594	2	T12488	hypothetical prote
23	344	15.4	337	2	S60520	finger protein ZNF
24	344	15.4	580	2	A37107	spermatogenesis pr
25	343	15.4	686	2	A34612	zinc finger protei
26	342	15.3	488	2	S47072	finger protein HZF
27	341	15.3	336	2	S06578	finger protein (cl
28	341	15.3	347	2	S00549	developmental cont
29	340	15.3	339	2	JC1442	transcription fact
30	339	15.2	411	2	S10245	finger protein, te
31	338	15.2	247	2	S47070	finger protein HZF
32	338	15.2	383	2	C32891	finger protein 9,
33	337.5	15.1	728	2	A48830	probable transcrip
34	337	15.1	469	2	I38600	zinc finger protei
35	337	15.1	595	2	JC7779	Krueppel-associate
36	336	15.1	553	2	S22954	finger protein zfp
37	335.5	15.1	428	2	S03677	finger protein (cl
38	334.5	15.0	280	2	S06572	finger protein (cl
39	332	14.9	536	2	S06548	finger protein (cl
40	332	14.9	701	2	T14757	hypothetical prote
41	331	14.8	794	2	S59069	Z13 protein - mous
42	330.5	14.8	732	2	S47073	finger protein HZF
43	329.5	14.8	496	2	T08674	probable finger pr
44	329	14.8	589	2	I38598	zinc finger protei
45	329	14.8	728	2	A54603	transcription fact

ALIGNMENTS

RESULT 1
S20050
transcription factor IIIA - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein P9677.9; protein YPR186c; TFC2 protein
C;Species: Saccharomyces cerevisiae
C;Date: 16-Sep-1992 #sequence_revision 16-Sep-1992 #text_change 01-Dec-2000
C;Accession: S20050; A44086; S58816
R;Archambault, J.; Milne, C.A.; Schappert, K.T.; Baum, B.; Friesen, J.D.; Segall, J.
J. Biol. Chem. 267, 3282-3288, 1992
A;Title: The deduced sequence of the transcription factor TFIIIA from Saccharomyces cere
A;Reference number: S20050; MUID:92147684; PMID:1737784
A;Accession: S20050
A;Molecule type: DNA
A;Residues: 1-429 <ARC>
A;Cross-references: EMBL:M80611; NID:g172902; PIDN:AAB08014.1; PID:g172903
R;Woychik, N.A.; Young, R.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 3999-4003, 1992
A;Title: Genes encoding transcription factor IIIA and the RNA polymerase common subunit
A;Reference number: A44086; MUID:92237295; PMID:1570325
A;Accession: A44086
A;Molecule type: DNA
A;Residues: 1-429 <WOY>
A;Cross-references: GB:M90638
R;Miller, N.
submitted to the EMBL Data Library, April 1995
A;Description: The sequence of S. cerevisiae cosmid 9677.
A;Reference number: S58816
A;Accession: S58816
A;Molecule type: DNA
A;Residues: 1-429 <MIL>
A;Cross-references: EMBL:U25841; NID:g786295; PID:g786305; MIPS:YPR186c
C;Genetics:
A;Gene: SGD:PZF1; TFIIIA; TFC2
A;Cross-references: SGD:S0006390; MIPS:YPR186c
A;Map position: 16R
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: DNA binding; nucleus; transcription factor; zinc finger

Query Match 37.6%; Score 837.5; DB 2; Length 429;
Best Local Similarity 43.7%; Pred. No. 1.2e-43;
Matches 174; Conservative 62; Mismatches 135; Indels 27; Gaps 11;

```
Qy          1 MSESDETKSISSLIS--SSSSSRPKKYICTYEGCDKAYNRPSLLEQH-LRTHSNDRPYKC 57
              :| |: ::|::|| |  ||||:||| | | |:|||||: |||:| :| |  |   | ::|
Db         23 ISRSESSESLNSLTSTRSSSSNRPKTYFCDYDGCDKAFTRPSILTEHQLSVHQGLRAFQC 82

Qy         58 TVDDCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHTKSFKCTFEN 117
                | | |:| :||||| |: :||: ||| || ||||| :|| |||||:|||||| |  |
Db         83 --DKCAKSFVKKSHLERHLYTHSDTKPFQCSYCGKGVTTRQQLKRHEVTHTKSFICPEEG 140

Qy        118 CQEAFYKHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGG--SPAYQCDHPG 175
                |  ||||  || ||||||   |||  ||| | || :|  |  |||       |||   |
```

```
Db        141 CNLRFYKHPQLRAHILSVHLHKLTCPHCNKSFQRPYRLRNHISKHHDPEVENPYQCTFAG 200

Qy        176 CFKNFQTWSVLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDV 235

Db        201 CCKEFRIWSQLQSHIKNDHPKLKCPICSKPCVGENGLQMHMIIHDDSLVTKNWKCHICPD 260

Qy        236 GKFAKKNELVEHYNIFH-DGNIPDDL-LKETEVKKLENLLDQGSKLNNLHELETEKLKVE 293

Db        261 MSFSRKHDLLTHYGSIHTEEDIPLELKYKISDIQQL--VQDHGVQLGN------SKHSNE 312

Qy        294 EDEEDEEDSLDEKRSDVRSDSMSAQRSIKSFTASLE-GSKSVSKLISNSGKKINCPKNNC 352

Db        313 QDEEKISNRLRKRRKLTENNNVEFLQNEVDLEKRLESGENGLNLLLNTVGRKYRCFYNNC 372

Qy        353 DRMFSREYDLRRHL---KWHDDNLQRIESFLNSIEKEE 387

Db        373 SRTFKTKEKYEKHIDKHKVHELKLKILQ------EKEE 404


RESULT 2
T37676
zinc-finger protein - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 01-Dec-2000
C;Accession: T37676
R;McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Davis, P.; Churcher, C.M.
submitted to the EMBL Data Library, October 1999
A;Reference number: Z21736
A;Accession: T37676
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-374 <MCD>
A;Cross-references: EMBL:AL132675; PIDN:CAB59689.1; GSPDB:GN00066; SPDB:SPAC144.09c
A;Experimental source: strain 972h-; cosmid c144
C;Genetics:
A;Gene: SPDB:SPAC144.09c
A;Map position: 1
A;Introns: 50/3
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

  Query Match             22.2%;  Score 495;  DB 2;  Length 374;
  Best Local Similarity   32.8%;  Pred. No. 5.7e-23;
  Matches  123;  Conservative   58;  Mismatches  160;  Indels   34;  Gaps   13;

Qy         15 SSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLET 74

Db         13 SKNLRSAKKIFHCPYEECGKKYSRPSLLEQHLRTHSNERPFVCDYTGCSKAFYRKSHLKI 72

Qy         75 HIVSHSEKKPFHCSV--CGKGVNSRQHLKRHEITH--TKSFKCTFENCQEAFYKHQSLRH 130

Db         73 HKRCHTNVKPFSCHYDGCDAQFYTQQHLERHIEVHRKPKPYACTWEGCDECFSKHQQLRS 132

Qy        131 HILSVHEKTL----TCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGC--FKNFQTWS 184

Db        133 HISACHTHLLPYPCTYQDCELRFATKQKLQNHVNRAHEKIISYSCPHESCVGHEGFEKWS 192

Qy        185 VLQFHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDV----GKFAK 240

Db        193 QLQNHIREAHVP-SCSICGRQFKTAAHLRHHVVLH--QTTLEERKTYHCPMEGCKKSFTR 249

Qy        241 KNELVEHYNIFHDGNIP---DDLLKETEVK-KLENLLDQGSKLNNLHELETEKLKVEEDE 296

Db        250 SSALKKHISVIHEGNMAFHCDSCGTKFGYKHMLQRHLERGT-CKKAHKPYINECGIKHDG 308

Qy        297 EDEEDSLDEKRSDVRSDSMS--AQRSIKSFTASLEGSKSVSKLISNSGKKINCPKNNCDR 354

Db        309 IEGVAIHDQKEKELSSNLVSDVAKKIINEVTG--HGYK--------EAREYSCSFPECNY 358

Qy        355 MFSREYDLRRHLKWH 369

Db        359 RFKRLYDMHRHLNSH 373


RESULT 3
S26823
zinc finger protein ZNF43 - human
N;Alternate names: zinc finger protein kox27
C;Species: Homo sapiens (man)
C;Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
C;Accession: S26823; I37967; S10416
R;Lovering, R.; Trowsdale, J.
Nucleic Acids Res. 19, 2921-2928, 1991
A;Title: A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell li
A;Reference number: S26823; MUID:91279444; PMID:1711675
A;Accession: S26823
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-803 <LOV>
A;Cross-references: EMBL:X59244; NID:g38031; PIDN:CAA41932.1; PID:g38032
R;Thiesen, H.J.
New Biol. 2, 363-374, 1990
A;Title: Multiple genes encoding zinc finger domains are expressed in human T cells.
A;Reference number: I37949; MUID:91145339; PMID:2288909
A;Accession: I37967
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 476-531 <THI>
A;Cross-references: EMBL:X52358; NID:g34160; PIDN:CAA36584.1; PID:g930090
C;Genetics:
A;Gene: GDB:ZNF43; HTF6
A;Cross-references: GDB:128653
A;Map position: 19p13.1-19p12
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: DNA binding; zinc finger

  Query Match             17.9%;  Score 400;  DB 2;  Length 803;
  Best Local Similarity   31.9%;  Pred. No. 7.5e-17;
  Matches  121;  Conservative   53;  Mismatches  171;  Indels   34;  Gaps   19;

Qy          2 SESDETKSISSLISSSSS--SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTV 59

Db        367 TECGEAFSRSSNLTKHKKIHTEKKPYKC--EECGKAFKWSSKLTEHKLTHTGEKPYKC-- 422

Qy         60 DDCDKAFFRKSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFEN 117

Db        423 EECGKAFNWPSTLTKHNRIHTGEKPYKCEVCGKAFNQFSNLTTHKRIHTAEKPYKC--EE 480

Qy        118 CQEAFYKHQSL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGC 176

Db        481 CGKAFSRSSNLTKHKKIHIEKKPYKCEECGKAFKWSSKLTEHKITHTGEKP-YKCEE--C 537

Qy        177 FKNFQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCD 234

Db        538 GKAFNHFSILTKHKRIHTGEKPYKCEECGKAFTQSSNLTTHKKIH---TGEKFYKCEEC- 593

Qy        235 VGK-FAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVE 293

Db        594 -GKAFTQSSNLTTHKKI-HTGGKP---YKCEECGKAFNQFSTLTKHKIIH-TEEKPYKCE 647

Qy        294 E-DEEDEEDSLDEKRSDVRSDS--MSAQRSIKSFTASLEGSKSVSKLISNSGKKINCPKN 350

Db        648 ECGKAFKWSSTLTKHKIIHTGEKPYKCEECGKAF--KLSSTLSTHKIIHTGEKPYKCEK- 704

Qy        351 NCDRMFSREYDLRRHLKWH 369

Db        705 -CGKAFNRPSNLIEHKKIH 722


RESULT 4
G02075
transcription repressor zinc finger protein 85 - human
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 01-Dec-2000
C;Accession: G02075
R;Poncelet, D.A.
submitted to the EMBL Data Library, September 1995
```

```
A;Reference number: G09169
A;Accession: G02075
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-595 <PON>
A;Cross-references: EMBL:U35376; NID:g1017721; PIDN:AAA79179.1; PID:g1017722
C;Genetics:
A;Gene: GDB:ZNF85
A;Cross-references: GDB:132279
A;Map position: 19p12-19p12
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

  Query Match             17.8%;  Score 396.5;  DB 2;  Length 595;
  Best Local Similarity   31.7%;  Pred. No. 8.8e-17;
  Matches  120;  Conservative   48;  Mismatches  160;  Indels   51;  Gaps   17;

Qy         10 ISSLISSSS-SSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFR 68
Db        186 ISCLTEHSRIHTRVNFYKC--EECGKAFNWSSTLTKHKRIHTGEKPYKC--EECGKAFNQ 241

Qy         69 KSHLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQ 126
Db        242 SSNLIKHKKIHTGEKPYKCEECGKAFNRFSTLTTHKIIHTGEKPYKC--KECGKAFNRSS 299

Qy        127 SL-RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSV 185
Db        300 TLTTHRKIHTGEKPYKCEECGKAFKQSSNLTTHKIIHTGEKP-YKCKK--CGKAFNQSAH 356

Qy        186 LQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNE 243
Db        357 LTTHEVIHTGEKPYKCEKCGKAFNHFSHLTTHKIIH---TGEKPYKCKEC--GKAFKHSS 411

Qy        244 LVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSL 303
Db        412 TLTKHKIIHTGE------KPYKSKECEKAFNQSSKL-------TEHKKIHTGEKPYE--- 455

Qy        304 DEKRSDVRSDSMSAQRSIKSFTASL--------EGSK-----SVSKLISNSGKKINCPKN 350
Db        456 CEKCGKAFNQSSNLTRHKKSHTEEKPYKCEECGKGFKWPSTLTIHKIIHTGEKPYKC--E 513

Qy        351 NCDRMFSREYDLRRHLKWH 369
Db        514 ECGKAFNQSSKLTKHKKIH 532


RESULT 5
I38937
DNA/RNA-binding protein - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jul-2000
C;Accession: I38937
R;Drew, P.D.; Nagle, J.W.; Canning, R.D.; Ozato, K.; Biddison, W.E.; Becker, K.G.
Gene 159, 215-218, 1995
A;Title: Cloning and expression analysis of a human cDNA homologous to Xenopus TFIIIA.
A;Reference number: I38937; MUID:95347600; PMID:7622052
A;Accession: I38937
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-363 <RES>
A;Cross-references: EMBL:U20272; NID:g644870; PIDN:AAA75623.1; PID:g644871
C;Superfamily: transcription factor IIIA

  Query Match             17.7%;  Score 395.5;  DB 2;  Length 363;
  Best Local Similarity   31.1%;  Pred. No. 5.9e-17;
  Matches   99;  Conservative   54;  Mismatches   96;  Indels   69;  Gaps   16;

Qy          8 KSISSL------ISSSSSSR--------PKKYICTYEGCDKAYNRPSLLEQHLRTHSNDR 53
Db          7 ESVSSLTIADAFIAAGESSAPTPPRPALPRRFICSFPDCSANYSKAWKLDAHLCKHTGER 66

Qy         54 PYKCTVDDCDKAFFRKSHLETHIVSHSEKKPFHCSV--CGKGVNSRQHLKRH-EITH--- 107
Db         67 PFVCDYEGCGKAFIRDYHLSRHILTHTGEKPFVCAANGCDQKFNTKSNLKKHFERKHENQ 126

Qy        108 TKSFKCTFENCQEAFYKHQSLR-HHILSVHEKTLTCKQ--CNKVFTRPSKLAQHKLKHHG 164
Db        127 QKQYICSFEDCKKTFKKHQQLKIHQCQHTNEPLFKCTQEGCGKHFASPSKLKRHAKAHEG 186

Qy        165 GSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK---------------------------- 196
Db        187 ----YVC-QKGCSFVAKTWTELLKHVRETHKEEILCEVCRKTFKRKDYLKQHMKTHAPER 241

Qy        197 --LKCPK--CGKGCVGKKGLSSHMLS-HDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNI 250
Db        242 DVCRCPREGCGRTYTTVFNLQSHILSFHEES---RPFVCEHAGCGKTFAMKQSLTRH-AV 297

Qy        251 FHDGNIPDDLLKETEVKK 268
Db        298 VHD---PDKKKMKLKVKK 312


RESULT 6
S35305
zinc finger protein ZNF91 - human
C;Species: Homo sapiens (man)
C;Date: 31-Dec-1993 #sequence_revision 02-Jun-1994 #text_change 05-Nov-1999
C;Accession: S35305
R;Bellefroid, E.J.; Marine, J.C.; Ried, T.; Lecocq, P.J.; Riviere, M.; Amemiya, C.; Ponce
EMBO J. 12, 1363-1374, 1993
A;Title: Clustered organization of homologous KRAB zinc-finger genes with enhanced expres
A;Reference number: S35305; MUID:93223677; PMID:8467795
A;Accession: S35305
A;Molecule type: mRNA
A;Residues: 1-1191 <BEL>
A;Cross-references: EMBL:L11672; NID:g186773; PIDN:AAA59469.1; PID:g186774
A;Note: the authors translated the codon GCA for residue 750 as Thr and GCT for residue 8
C;Genetics:
A;Gene: GDB:ZNF91; HPF7; HTF10
A;Cross-references: GDB:132284
A;Map position: 19p12-19p12
C;Keywords: DNA binding; zinc finger

  Query Match             17.7%;  Score 393.5;  DB 2;  Length 1191;
  Best Local Similarity   35.8%;  Pred. No. 2.8e-16;
  Matches   98;  Conservative   40;  Mismatches  109;  Indels   27;  Gaps   12;

Qy         20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
Db        905 TREKPYKC--EECGKAFSQPSHLTTHKRMHTGEKPYKC--EECGKAFSQSSTLTTHKIIH 960

Qy         80 SEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-RHHILSVH 136
Db        961 TGEKPYKCEECGKAFRKSSTLTEHKIIHTGEKPYKC--EECGKAFSQSSTLTRHTRMHTG 1018

Qy        137 EKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFHIKQSHPK 196
Db       1019 EKPYKCEECGKAFNRSSKLTTHKIIHTGEKP-YKCEE--CGKAFISSSTLNGH-KRIHTR 1074

Qy        197 ---LKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHD 253
Db       1075 EKPYKCEECGKAFSQSSTLTRHKRLH---TGEKPYKCGEC--GKAFKESSALTKHKIIHT 1129

Qy        254 GNIPDDLLKETEVKKLENLLDQGSKLNNLHELET 287
Db       1130 GE------KPYKCEKCCKAFNQSSILTNHKKIHT 1157


RESULT 7
S00647
finger protein - African clawed frog
C;Species: Xenopus laevis (African clawed frog)
C;Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 31-Dec-1993
C;Accession: S00647
R;Ruiz i Altaba, A.; Perry-O'Keefe, H.; Melton, D.A.
EMBO J. 6, 3065-3070, 1987
A;Title: Xfin: an embryonic gene encoding a multifingered protein in Xenopus.
```

A;Reference number: S00647; MUID:88082679; PMID:2826129
A;Accession: S00647
A;Molecule type: mRNA
A;Residues: 1-1350 <RUI>
A;Cross-references: EMBL:X06021
A;Note: it is uncertain whether Met-1, Met-11, Met-16, Met-38, or Met-39 is the initiato
C;Genetics:
A;Gene: fin
C;Keywords: DNA binding; zinc finger

Query Match 16.8%; Score 375.5; DB 2; Length 1350;
Best Local Similarity 27.3%; Pred. No. 4e-15;
Matches 119; Conservative 65; Mismatches 171; Indels 81; Gaps 20;

Qy 11 SSLISSSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKS 70
Db 340 SFLNHQQTHSREKPYLCSH--CNKGFIQNSDLVKHFRTHTGERPYQCA--ECHKGFIQKS 395

Qy 71 HLETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL 128
Db 396 DLVKHLRTHTGEKPFKCSHCDKKFTERSALAKHQRTHTGEKPYKCS--DCGKEFTQRSNL 453

Qy 129 -RHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKH--------HGGSPAYQCDHPGCFKN 179
Db 454 ILHQRIHTGERPYKCTLCDRTFIQNSDLVKHQKVHANLPLSDPHTANSPHKCSK--CDLT 511

Qy 180 FQTWSVLQFHIK--QSHPKLKCPKCGKGCVGKKGLSSHMLSH----------DD 221
Db 512 FSHWSTFMKHSKLHSGEKKFQCAECKKGFTQKSDLVKHIRVHTGEKPFKCLLCKKSFSQN 571

Qy 222 STMIKIW---------TCDYCDVGKFAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENL 272
Db 572 SDLHKHWRIHTGEKPFPCYTCD-KSFTERSALIKHHRT-HTGERPH------KCSVCQKG 623

Qy 273 LDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDSMSAQR----------S 320
Db 624 FIQKSAL-------TKHSRTHTGEKPYPCTQCGKSFIQNSDLVKHQRIHTGEKPYHCTEC 676

Qy 321 IKSFTASLEGSKSVSKLISNSGKK-INCPKNNCDRMFSREYDLRRHLKWHD-DNLQRIES 378
Db 677 NKRFT---EGSSLVKHRRTHSGEKPYRCPQ--CEKTFIQSSDLVKHLVVHNGENPPAATA 731

Qy 379 FLN-SIEKEETPEGEP 393
Db 732 FHEILIRRENLTRSEP 747

RESULT 8
I48668
zinc finger protein 51 - mouse
N;Alternate names: finger protein zfec12
C;Species: Mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 08-Dec-2000
C;Accession: I48668; A40984
R;Burke, P.S.; Don, J.; Wolgemuth, D.J.
Mamm. Genome 5, 387-389, 1994
A;Title: Zfp-51, a murine zinc finger encoding gene mapping to the t-complex region of c
A;Reference number: I48668; MUID:94319090; PMID:8043957
A;Accession: I48668
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-710 <RES>
A;Cross-references: EMBL:X74855; NID:g488832; PIDN:CAA52847.1; PID:g488833
R;Crossley, P.H.; Little, P.F.R.
Proc. Natl. Acad. Sci. U.S.A. 88, 7923-7927, 1991
A;Title: A cluster of related zinc finger protein genes is deleted in the mouse embryoni
A;Reference number: A40984; MUID:91376058; PMID:1680234
A;Accession: A40984
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 672-710 <CRO>
A;Cross-references: GB:M74235
C;Genetics:
A;Gene: Zfp-51
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: DNA binding; zinc finger

Query Match 16.8%; Score 374.5; DB 2; Length 710;
Best Local Similarity 29.6%; Pred. No. 2.3e-15;
Matches 104; Conservative 54; Mismatches 140; Indels 53; Gaps 15;

Qy 25 YICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSHSEKKP 84
Db 311 YEC--KECGKSFIELSHLKRHYRIHTGEKPYKCEI--CDKSFTTTTTLKTHQKIHTGEKP 366

Qy 85 FHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHHILS-VHEKTLT 141
Db 367 YKCRECDKSFIHSSHLRRHQNVHTGERPYRC--KECDKSFHESATLREHEKSHTGEKTYK 424

Qy 142 CKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSVLQFH--IKQSHPKLKC 199
Db 425 CRECDKSFTQRAYLRNHHNRVHTGERPYECKE--CGKSFTTCSTLRIHQTIHTGKKPYKC 482

Qy 200 PKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEHYNIFHDGNIPDD 259
Db 483 IECGKSFTNNSYLRTHQKVHSGE---KPYRCKECD-KSFTSCSTLKAHQSI-HTGEKPYK 537

Qy 260 LLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEEDSLDEKRSDVRSDSMSAQR 319
Db 538 CI------ECDKSFTQVSHLRTHQRVHTGERPFRCTECD--------KSFIRSSHLREHQ 583

Qy 320 SIKSFTASLEGSKSVSKLISNSGKK-INCPKNNCDRMFSREYDLRRHLKWH 369
Db 584 KI------------------HSGEKPYKC--RDCDISFSQISNLRRHQKLH 614

RESULT 9
A54661
zinc finger protein ZNF41 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 01-Dec-2000
C;Accession: A54661; I54227
R;Franze, A.; Archidiacono, N.; Rocchi, M.; Marino, M.; Grimaldi, G.
Genomics 9, 728-736, 1991
A;Title: Isolation and expression analysis of a human zinc finger gene (ZNF41) located o
A;Reference number: A54661; MUID:91244317; PMID:2037297
A;Accession: A54661
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-542 <FRA>
A;Cross-references: GB:M36548; GB:M92443; NID:g340443; PIDN:AAA61312.1; PID:g340444
C;Genetics:
A;Gene: GDB:ZNF41
A;Cross-references: GDB:125865; OMIM:314995
A;Map position: Xp11.3-Xp11.23
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 16.7%; Score 373; DB 2; Length 542;
Best Local Similarity 29.0%; Pred. No. 2.1e-15;
Matches 115; Conservative 59; Mismatches 156; Indels 66; Gaps 21;

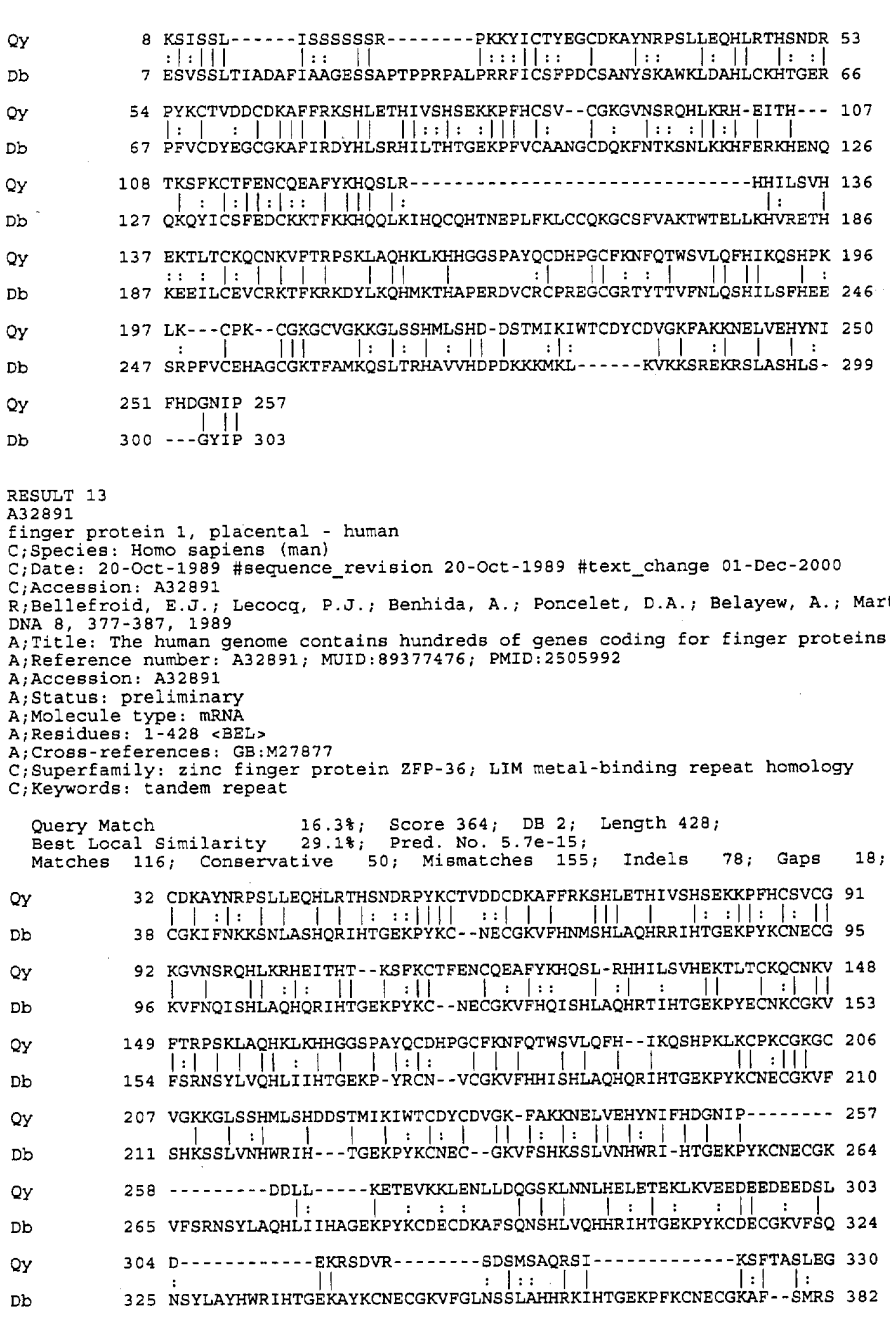
Qy 14 ISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLE 73
Db 163 IHQKTHTGEKHYEC--NECGKAFTRKSALRMHQRIHTGEKPYVCA--DCGKAFIQKSHFN 218

Qy 74 THIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSLRHH 131
Db 219 THQRIHTGEKPYECSDCGKSFTKKSQLHVHQRIHTGEKPYICT--ECGKVFTHRTN---- 272

Qy 132 ILSVHEKTLT------CKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWSV 185
Db 273 -LTTHQKTHTGEKPYMCAECGKAFTDQSNLIKHQKTHTGEKP-YKCN--GCGKAFIWKSR 328

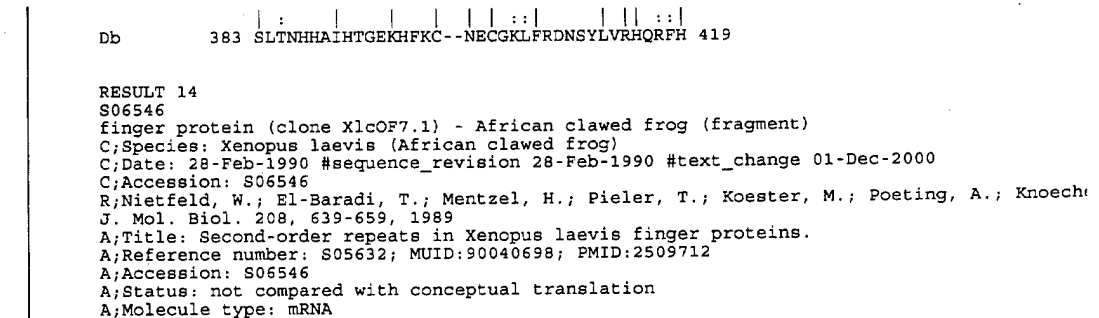
Qy 186 LQFHIKQSH---PKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKK 241
Db 329 LKIH-QKSHIGERHYECKDCGKAFIQKSTLSVHQRIH---TGEKPYVCPEC--GKAFIQK 382

```
Qy        242 NELVEHYNIFHDGNIPDD-------LLKETEVKKLENLLDQGSKLNNLHE---------- 284
              :  : |: | | |   | :          |:::: ::   :  | | |   |
Db        383 SHFIAHHRI-HTGEKPYECSDCGKCFTKKSQL-RVHQKIHTGEKPNICAECGKAFTDRSN 440

Qy        285 -LETEKLKVEEDEEDEEDS----LDEKRSDVRSDSMSAQRSI------KSFTASLEGSKS 333
               :  :|:    |   :  |        : | ::   | : :|         |:|    : : |
Db        441 LITHQKIHTREKPYECGDCGKTFTWKSRLNIHQKSHTGERHYECSKCGKAFIQ--KATLS 498

Qy        334 VSKLISNSGKKINCPKNNCDRMFSREYDLRRHLKWH 369
              : ::|     |   |    | : |:   :| :| | |
Db        499 MHQIIHTGKKPYAC--TECQKAFTDRSNLIKHQKMH 532


RESULT 10
B32891
finger protein 2, placental - human
C;Species: Homo sapiens (man)
C;Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 01-Dec-2000
C;Accession: B32891
R;Bellefroid, E.J.; Lecocq, P.J.; Benhida, A.; Poncelet, D.A.; Belayew, A.; Martial, J.A
DNA 8, 377-387, 1989
A;Title: The human genome contains hundreds of genes coding for finger proteins of the K
A;Reference number: A32891; MUID:89377476; PMID:2505992
A;Accession: B32891
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-651 <BEL>
A;Cross-references: GB:M27878
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: tandem repeat

  Query Match             16.7%;  Score 373;  DB 2;  Length 651;
  Best Local Similarity   27.4%;  Pred. No. 2.6e-15;
  Matches  128;  Conservative   57;  Mismatches  146;  Indels  136;  Gaps   24;

Qy         23 KKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH--- 79
              | | |:    | ||::: | | |  | |||: ::||:|   :| ||| ||||| :|  :|
Db        174 KPYNCSQ--CGKAFSQKSQLTSHQRTHTGEKPYEC--GECGKAFSRKSHLISHWRTHTGE 229

Qy         80 -----------SEK------------KPFHCSVCGKGVNSRQHLKRHEITH--TKSFK 112
                         |||            ||| |  |||  : :  |  |  ||  || | |
Db        230 KPYGCNECGRAFSEKSNLINHQRIHTGEKPFECRECGKAFSRKSQLVTHHRTHTGTKPFG 289

Qy        113 CTFENCQEAFY-KHQSLRHHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSP---- 167
              |:  :|::||: | : :||  :   ||    | :| | |    | |  |:  | |  |
Db        290 CS--DCRKAFFEKSELIRHQTIHTGEKPYECSECRKAFRERSSLINHQRTHTGEKPHGCI 347

Qy        168 ------------AYQCDHPG--------CFKNFQTWSVLQFHIKQSHPKLK---CPKCG 203
                        ::|  | |          | | |   | | | :::|   |    | :||
Db        348 QCGKAFSQKSHLISHQMTHTGEKPFICSKCGKAFSRKSQLVRH-QRTHTGEKPYECSECG 406

Qy        204 KGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGK-FAKKNELVEHYNIFHDGNIPDDLL- 261
              |    |  |::|   |   |  | : |  |  || | :|: |: |     | |  | :
Db        407 KAFSEKLSLTNHQRIH---TGEKPYVCSEC--GKAFCQKSHLISHQRT-HTGEKPYECSE 460

Qy        262 --------------------KETEVKKLENLLDQGSKLNNLHELETEKLKVEEDEEDEE 300
                                 | | : |   | |:|| |:       ::   |:  |
Db        461 CGKAFGEKSSLATHQRTHTGEKPYECRDCEKAFSQKSQLNT-HQ------RIHTGEKPYE 513

Qy        301 DSLDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLISNSGKK-INCPKNNCDRMFSRE 359
              ||  |                | | |: :  | :::|:|   |  | | : |  :
Db        514 CSLCRK------------------AFFEKSELIRHLRTHTGEKPYEC--NECRKAFREK 552

Qy        360 YDLRRHLKWHDDNLQRIESFLNSIEKEETPEGEPLVKKARMDLLPNE 406
                | |         |||  :     |  |   |:   :|:   |:|::
Db        553 SSLINH--------QRIHTGEKPFECSEC--GKAFSRKSH--LIPHQ 587


RESULT 11
JN0533
finger protein pMLZ-4 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 01-Dec-2000
C;Accession: JN0533
R;Brady, J.P.; Piatigorsky, J.
Gene 124, 207-214, 1993
A;Title: Cloning and characterization of a novel zinc-finger protein encoding cDNA from t
A;Reference number: JN0533; MUID:93185925; PMID:8444344
A;Accession: JN0533
A;Molecule type: mRNA
A;Residues: 1-393 <BRA>
A;Cross-references: GB:M98502; NID:g200406; PIDN:AAA39949.1; PID:g200407
A;Experimental source: eye lens
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: DNA binding; eye lens; zinc finger
F;1-62/Region: acidic
F;63-391/Region: zinc finger

  Query Match             16.7%;  Score 372;  DB 2;  Length 393;
  Best Local Similarity   29.4%;  Pred. No. 1.7e-15;
  Matches  110;  Conservative   57;  Mismatches  163;  Indels   44;  Gaps   17;

Qy         13 LISSSSSSRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHL 72
              |:   : :  : :||:   | ||::: | | :| :||: ||||||   :| | | | |||
Db         49 LVHKEAHAGIRYHICSQ--CGKAFSQISDLNRHQKTHTGDRPYKCY--ECGKGFSRSSHL 104

Qy         73 ETHIVSHSEKKPFHCSVCGKGVNSRQHLKRHEITHT--KSFKCTFENCQEAFYKHQSL-R 129
                |. :|: ::|: |: |||       || :|:  ||  |  |||    | :|      | :
Db        105 IQHQRTHTGERPYDCNECGKSFGRSSHLIQHQTIHTGEKPHKCT--ECAKASAASPHLIQ 162

Qy        130 HHILSVHEKTLTCKQCNKVFTRPSKLAQHKLKHHGGSPAYQCDHPGCFKNFQTWS--VLQ 187
              |     ||   |::| | |:| | ||||:  | |  | |:|    | : |    |  :
Db        163 HQRTHSGEKPYECEECGKSFSRSSHLAQHQRTHTGEKP-YECHE--CGRGFSERSDLIKH 219

Qy        188 FHIKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKIWTCDYCDVGKFAKKNELVEH 247
              : :       || :|||       |  |  :|   |  | : |: |    |:: : ||:|
Db        220 YRVHTGERPYKCDECGKNFSQNSDLVRHRRAH---TGEKPYHCNECG-ENFSRISHLVQH 275

Qy        248 YNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELETEKLKVEEDE-----EDEEDS 302
              | |         | |          : | |     :: | :    | :|       :  |
Db        276 QRT-HTGE------KPYECTACGKSFSRSSHLITHQKIHTGEKPYECNECWRSFGERSDL 328

Qy        303 LDEKRSDVRSDSMSAQRSIKSFTASLEGSKSVSKLIS----NSGKK-INCPKNNCDRMFS 357
              :  :|:           :  | ||| |        | ||:    ::|:|    |    ||: ||
Db        329 IKHQRTHTGEKPYECVQCGKGFTQS-------SNLITHQRVHTGEKPYEC--TECDKSFS 379

Qy        358 REYDLRRHLKWHDD 371
              |   | :| : | |
Db        380 RSSALIKHKRVHTD 393


RESULT 12
G01496
transcription factor IIIA - human (fragment)
N;Alternate names: TFIIIA
C;Species: Homo sapiens (man)
C;Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 05-Nov-1999
C;Accession: G01496
R;Becker, K.G.
submitted to the EMBL Data Library, August 1994
A;Reference number: G07520
A;Accession: G01496
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-338 <BEC>
A;Cross-references: EMBL:U14134; NID:g551534; PIDN:AAA21873.1; PID:g551535
C;Genetics:
A;Gene: GDB:GTF3A; TFIIIA
A;Cross-references: GDB:434744; OMIM:600860
A;Map position: 13q12.3-13q13.1
C;Superfamily: transcription factor IIIA

  Query Match             16.4%;  Score 364.5;  DB 2;  Length 338;
  Best Local Similarity   29.0%;  Pred. No. 4.1e-15;
  Matches   89;  Conservative   49;  Mismatches  102;  Indels   67;  Gaps   11;
```



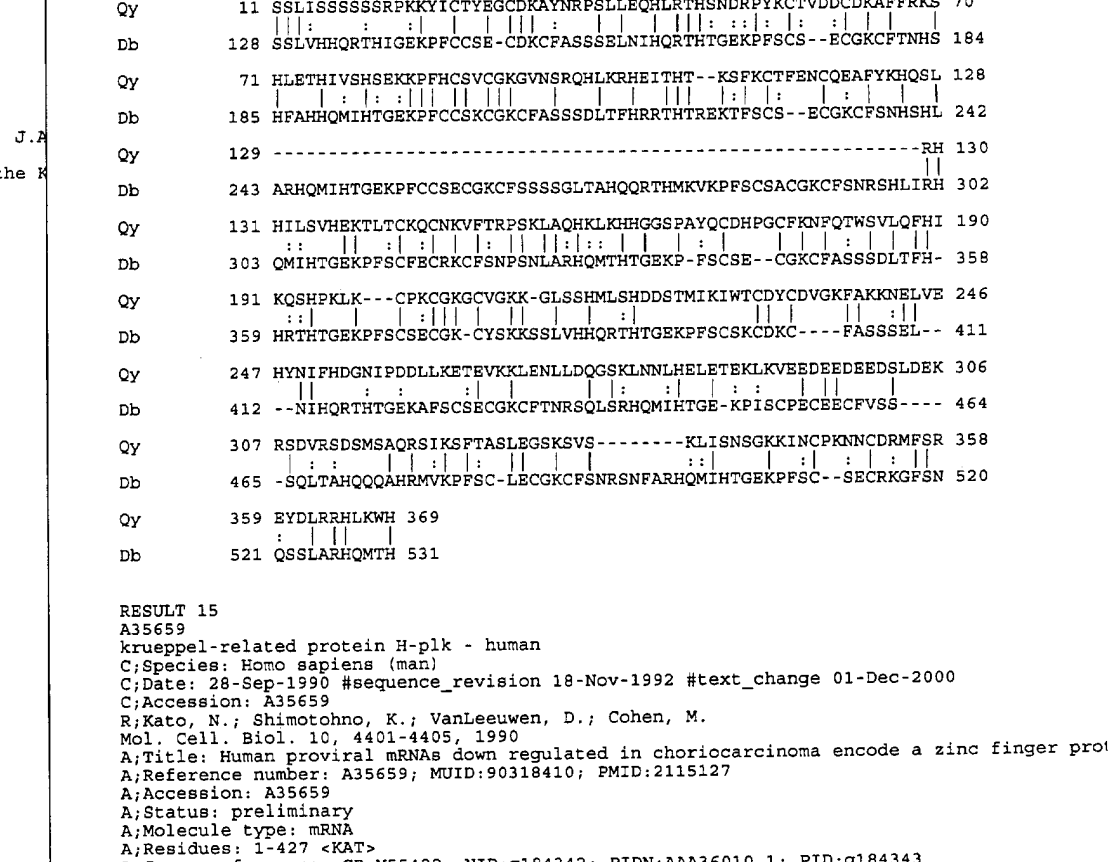
RESULT 13
A32891
finger protein 1, placental - human
C;Species: Homo sapiens (man)
C;Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 01-Dec-2000
C;Accession: A32891
R;Bellefroid, E.J.; Lecocq, P.J.; Benhida, A.; Poncelet, D.A.; Belayew, A.; Martial, J.A
DNA 8, 377-387, 1989
A;Title: The human genome contains hundreds of genes coding for finger proteins of the K
A;Reference number: A32891; MUID:89377476; PMID:2505992
A;Accession: A32891
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-428 <BEL>
A;Cross-references: GB:M27877
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: tandem repeat

Query Match 16.3%; Score 364; DB 2; Length 428;
Best Local Similarity 29.1%; Pred. No. 5.7e-15;
Matches 116; Conservative 50; Mismatches 155; Indels 78; Gaps 18;



RESULT 14
S06546
finger protein (clone XlcOF7.1) - African clawed frog (fragment)
C;Species: Xenopus laevis (African clawed frog)
C;Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 01-Dec-2000
C;Accession: S06546
R;Nietfeld, W.; El-Baradi, T.; Mentzel, H.; Pieler, T.; Koester, M.; Poeting, A.; Knoechl
J. Mol. Biol. 208, 639-659, 1989
A;Title: Second-order repeats in Xenopus laevis finger proteins.
A;Reference number: S05632; MUID:90040698; PMID:2509712
A;Accession: S06546
A;Status: not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-615 <NIE>
C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
C;Keywords: DNA binding; zinc finger

Query Match 16.3%; Score 363.5; DB 2; Length 615;
Best Local Similarity 27.6%; Pred. No. 9.1e-15;
Matches 119; Conservative 48; Mismatches 165; Indels 99; Gaps 18;



RESULT 15
A35659
krueppel-related protein H-plk - human
C;Species: Homo sapiens (man)
C;Date: 28-Sep-1990 #sequence_revision 18-Nov-1992 #text_change 01-Dec-2000
C;Accession: A35659
R;Kato, N.; Shimotohno, K.; VanLeeuwen, D.; Cohen, M.
Mol. Cell. Biol. 10, 4401-4405, 1990
A;Title: Human proviral mRNAs down regulated in choriocarcinoma encode a zinc finger prot
A;Reference number: A35659; MUID:90318410; PMID:2115127
A;Accession: A35659
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-427 <KAT>
A;Cross-references: GB:M55422; NID:g184342; PIDN:AAA36010.1; PID:g184343

C;Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 16.1%; Score 359.5; DB 2; Length 427;
Best Local Similarity 32.0%; Pred. No. 1.1e-14;
Matches 99; Conservative 36; Mismatches 121; Indels 53; Gaps 13;

```
Qy          20 SRPKKYICTYEGCDKAYNRPSLLEQHLRTHSNDRPYKCTVDDCDKAFFRKSHLETHIVSH 79
               :|    | |   |    :|:|  | | :| | |: ::||||   :| ||| : |||  |    |
Db         132 TRVNFYKC--EAYGRAFNWSSTLNKHKRIHTGEKPYKC--KECGKAFNQTSHLIRHKRIH 187

Qy          80 SEKKPFHCSVCGKGVNSRQHLKRHEITHTKS--FKCTFENCQEAFYKHQSL-RHHILSVH 136
               :|:||: |  |||  |    |  | | ||    :||  | |  || :   |  | ::
Db         188 TEEKPYKCEECGKAFNQSSTLTTHNIIHTGEIPYKC--EKCVRAFNQASKLTEHKLIHTG 245

Qy         137 EKTLTCKQCNKVFTRPSKLAQHKLKH--------------------------HGGSPAY 169
               ||   |::| | | | ||| :||  |                          | |   |
Db         246 EKRYECEECGKAFNRSSKLTEHKYIHTGEKLYKCEECGKAFNQSSTLTTHKRIHSGEKPY 305

Qy         170 QCDHPGCFKNFQTWSVLQFH--IKQSHPKLKCPKCGKGCVGKKGLSSHMLSHDDSTMIKI 227
               :|:   | | |: :| |  |  |        || :|||       |: | : |   | |
Db         306 KCEE--CGKAFKQFSNLTDHKKIHTGEKPYKCEECGKAFNQLSNLTRHKVIH---TGEKP 360

Qy         228 WTCDYCDVGKFAKKNELVEHYNIFHDGNIPDDLLKETEVKKLENLLDQGSKLNNLHELET 287
               : |  |  ||   ::  :  : | | |  |    :  :|  |       |||:   :: |
Db         361 YKCGEC--GKAFNQSSALNTHKIIHTGENPHKCRESGKVFHL------SSKLSTCKKIHT 412

Qy         288 -EKL-KVEE 294
                ||| | ||
Db         413 GEKLYKCEE 421
```

Search completed: May 5, 2004, 15:23:24
Job time : 21 secs